

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 11:27:40 ; Search time 8363.76 Seconds
(without alignments)
1360.163 Million cell updates/sec

Title: US-09-422-999-17
Perfect score: 3013
Sequence: 1 gatccagcagagatgtgat.....ccacaatttcaaaatgccc 3013

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est35:*
80: gb_est36:*
81: gb_est37:*
82: gb_est38:*
83: em_est39:*
84: em_est40:*
85: em_est41:*
86: em_est42:*
87: gb_est43:*
88: gb_est44:*
89: gb_est45:*
90: gb_est46:*
91: gb_est47:*
92: em_est48:*
93: em_est49:*
94: em_est50:*
95: em_est51:*
96: em_est52:*
97: em_est53:*
98: em_est54:*
99: gb_est55:*
100: gb_est56:*
101: em_est57:*
102: gb_est58:*
103: gb_est59:*
104: gb_est60:*
105: gb_est61:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	541.4	18.0	675	64	AL119269	AL119269 DKFZ761M
2	527	17.5	600	42	AI099199	AI099199 uc40c07.y

```
3 402.6 13.4 409 64 AL119765 AL119765 DKFZp761D
4 387.2 12.9 446 62 R59283 R59283 Y97H01.r1
5 342.4 11.4 541 69 AU079207 AU079207 AU079207
6 325.8 10.8 411 64 AM047433 AM047433 UT-M-BH1-
7 315.2 10.5 361 33 AA388005 AA388005 vc85b01.r
8 312.8 10.4 317 41 AI003002 AI003002 am95g07.s
9 280.6 9.3 343 64 AL120018 AL120018 DKFZp761H
10 268 8.9 268 43 AL120513 AL120513 an04h10.x
11 263.2 8.7 450 61 AI842094 AI842094 UT-M-ANI-
12 239.8 8.0 277 48 AI608196 AI608196 vc85b01.y
13 220.4 7.3 441 45 AI386336 AI386336 mn97c11.y
14 215.2 7.1 540 62 AI909224 AI909224 IL-BT202-
15 163.6 5.4 444 69 AM139646 AM139646 UT-H-BI1-
16 158.8 5.3 583 35 AA557278 AA557278 n176a05.s
17 148.4 4.9 413 64 AM084109 AM084109 xc37c10.x
18 141.8 4.7 952 80 CNS014C0 CNS014C0 AL103966
19 141.4 4.7 957 78 CNS012W6 CNS012W6 AL102036
20 133.4 4.4 267 28 AA087876 AA087876 dn97c11.r
21 127 4.2 538 39 AA890583 AA890583 aJ93d03.s
22 112.4 3.7 365 23 R88026 R88026 ym87e02.r1
23 101.4 3.4 513 82 AI196680 AI196680 u166d02.x
24 100.6 3.3 536 82 AQ717541 AQ717541 HS_2150_A
25 91.4 3.0 276 22 R47818 R47818 yJ56c04.r1
26 90.2 3.0 444 61 AI844405 AI844405 UT-M-ALI-
27 89.8 3.0 504 62 AI876724 AI876724 uJ34d05.y
28 88.8 2.9 378 37 AA687472 AA687472 ns16h10.s
29 82.8 2.7 459 31 AA291192 AA291192 z847b08.r
30 82.8 2.7 484 30 AA236800 AA236800 z776h12.s
31 82.8 2.7 493 39 AA890211 AA890211 ak10c10.s
32 81.8 2.7 432 33 AA453497 AA453497 zx47c12.r
33 81.2 2.6 536 41 AI228074 AI228074 EST224769
34 79.4 2.6 593 61 AI830871 AI830871 wJ61c07.x
35 78.6 2.6 435 41 AI018543 AI018543 cu30e10.x
36 73.4 2.4 175 41 AI060273 AI060273 UT-R-C1-1
37 72.8 2.4 459 43 AI227485 AI227485 u110b12.y
38 72.4 2.4 530 42 AI137215 AI137215 UT-R-C2p-
39 72.2 2.4 454 38 AA815190 AA815190 oc04e12.s
40 67.4 2.2 408 42 AI137445 AI137445 UT-R-C2b-
41 66.8 2.2 300 30 AA236846 AA236846 z776h12.r
42 66.2 2.2 528 37 AA707689 AA707689 zh23d05.s
43 66 2.2 541 62 AI875340 AI875340 UK31d03.y
44 65.4 2.2 457 25 W15550 W15550 zb91h09.s1
45 64.4 2.1 569 49 AI663371 AI663371 UK32a11.y
```

ALIGNMENTS

```
RESULT 1
LOCUS AL119269 675 bp mRNA EST 27-SEP-1999
DEFINITION DKFZp761M0513_r1 761 (synonym: hamy2) Homo sapiens cDNA clone
ACCESSION AL119269
VERSION DKFZp761M0513.5, mRNA sequence.
KEYWORDS AL119269.1 GI:5925168
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 675)
AUTHORS Ansoerge,W., Benes,V., Krieger,S., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
EST (Ansoerge, Benes, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246636.
Contact: Ansoerge W
MIPS
Am Kioferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL within the cDNA sequencing consortium of the
```

```
FEATURES
source
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="DKFZp761M0513"
    /clone_1lb="761 (synonym: hamy2)"
    /tissue_type="amygdala"
    /dev_stage="adult"
    /lab_host="DH10B"
    /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"
Location/Qualifiers
    1..675
```

```
BASE COUNT 179 a 172 c 158 g 166 t
ORIGIN
```

```
Query Match 18.0%; Score 541.4; DB 64; Length 675;
Best Local Similarity 99.1%; Pred. No. 2.6e-143; Indels 4; Gaps 3;
Matches 576; Conservative 0; Mismatches 1;
```

```
QY 1 gatcagcgaagatgtgatataatcttcacgcgaaggaagtgtaagctttgaga 60
|||||
DB 95 GATCCAGCGAAGATGTGATATATCTTCAGCTGAGTGAAGAGTAAAGCTTTGAGA 154
61 aattcaccacaatctcctcatcagattgcttatgtgtattatgagaatctgga 120
|||||
DB 155 AATTTTACCCTTCCCTTCATCATGATTTGCTTATGTATGAAATCTGGAAA 214
121 aggaataaataattctcccaagtgata-tggaacaacggtatgctg-cctggcag 178
|||||
DB 215 AGGAATATACATTTTCGCGAGGTGATTTGGAACAACTGTAAGCTTGGCAG 274
179 ggtcttgatgttaagatcttgagacagcagtcacccaaggaatcgtgacatctga 238
|||||
DB 275 GGTCTTGGATGTTAAAGTATGACACGAGTACCAAGATGCTGTGACATCTGTA 334
239 cctctggaatgtgagcgccttggaagatcgaatcttggaacaacaccccgcatgca 298
|||||
DB 335 CCTTGGGAATTTGGAGCGCTTTGGAGAGTTCATTTCTTGGAACAACCCCGCATGCA 394
299 ccacgttaacaggaagacagtgacactctccgcacagcagaagaagacatcaaggcac 358
|||||
DB 395 CCATCGTTACGAGGAGACAGTGAACCTCTCGCATCAAGAGAAAGACTTCAAGGCCAC 454
359 tatggagaagaatctgacagatataatgacaggaacttgcctcctcttatgtattatgg 418
|||||
DB 455 TATGGAGAAATATCGACGATATATGCGAGACTTCTGCTCCTCTTATGTTATG 514
419 aaacgggcttaacaatgacaggaatctctgcaagaagagaacac--ctcatgaaactc 476
|||||
DB 515 AAACGGGCTCTTAACAATGACAGATTTCTGACAAAGAGAACACACCTTCATTGAACCTC 574
477 acgtctctctgctctgctgtaacacattacaaaggtccctcaagaagaatccccaagg 536
|||||
DB 575 ACCTTCCTCTTCTGCTCTGCTACACCATTAACCAAGTCCCTTCAAGAGATCTCAGAG 634
537 ctggaanaatttaccgaatgcatctctctcgagcaact 577
|||||
DB 635 CTGGAATAATTTTACGAATGCAATTCCTCTCGAGACCACT 675
```

```
RESULT 2
LOCUS AI099199 600 bp mRNA EST 20-AUG-1998
DEFINITION uc40c07.y1 Sugano mouse liver mla Mus musculus cDNA clone
IMAGE:1482628 5' similar to WP:TT20G5.5 CE00479 CAMP DEPENDENT
PROTEIN KINASE REGULATORY CHAIN + GUANINE NUCLEOTIDE RELEASING
FACTORS ;, mRNA sequence.
ACCESSION AI099199
VERSION AI099199.1 GI:3448724
```

	EST.
SOURCE	house mouse-
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 600)
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HMNI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Apr 18, 1995 this sequence version replaced gi:775315. Contact: Marra M/Mouse EST Project WashU-HMNI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:931184 Seq primer: custom primer used High quality sequence stop: 511. Location/Qualifiers 1..600 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1482828" /clone_lib="Sugano mouse liver mlia" /sex="female" /dev_stage="adult" /lab_host="DHI0B" /note="Organ: liver; Vector: pME18S-FL3; Site:1: DraIII (CACCTGTGTG); Site:2: DraIII (CACCATGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTTCCTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for 3' end sequencing: 5' end primer CTTCGTCTTAAGAACTGTGG and 3' primer CGACCTCGACCTCGACACA."
BASE COUNT	172 a 134 c 153 g 140 t 1 others
ORIGIN	.
Query Match	17.5%; Score 527; DB 42; Length 600;
Best Local Similarity	92.3%; Pred. No. 3.2e-139;
Matches 554; Conservative	0; Mismatches 46; Indels 0; Gaps 0;
Dn	844 ttatgatgaggtcccggaacacatgtctgtgtgtctacagatgggcccccagccccacat 903 1 TGATGAAGAATTCAAGGACACCATTGTGTGCTCTCAAGATGGGGCCCTGACGCCCACT 60
Oy	904 gagagtcatccttcgcgaaccacacctgcgcagaggaactgtgatgacctagatatcta 963 61 GAGAAATGATCCCTCGCAAACACACTGCGCAGAGAGATGTGGATGACCTTAGAGATTATCTA 120
Dn	964 tgaggagcttctcatattaaagaaccttatoccatcttttacacaagtgaaaagagatt 1023 121 CGACGAGCTCTCTCAATATAAAGCCCTTAACCATCTCTTACACACAGTAAACGGGAGTT 180
Oy	1024 agagagtgtcttcatttttgttagtcttcacgcacaaagagaggaactgtttttaaccaagg 1083 181 AGCGGATGTTCTCATATTTTGTAGTCTACGCGCAAGAGAGAACTGTGTGTTTAACAGGG 240

QY	1084	ggaagaaggtaccccttcggtacatattctttaaagaagtaagtgatgtacatttcg	1143
Db	241	GGAAAGAAAGGTACCTCTCGTGTGATCATATTCTTGAAGGATCCGTGAATGTAGCATTTATG	300
QY	1144	caaggtgtgtcttcgcacccctgcataagaagatgacttcgccaagttagactgtgaa	1203
Db	301	CAAGGTGTGTGTCTGCAACCCTCGACGAAGGATGACTTTGGCAAGTTAGCTTAAGTGA	360
QY	1204	tgtatgcccaagctgcgtccttctatcgtcttcggaagaatactgcatttcttaaggt	1263
Db	361	CGATGCTCCAAAGACGTGCCTCCATTTGTTCTTCGGGAAGATTAATTTGCTCACTTAAGT	420
QY	1264	agacaagaagatttcaaccgcatcccaagaagcgttgaaggcgaatacagtcacacttaa	1323
Db	421	CGACAGGAAGACCTTCATTCGGATTTCGAGGAGCTTGAGCGAATACGTCACACTTAA	480
QY	1324	agaacatgaccagaatgtcttgtgtgtcgtgaagaaggtcccaagcagaagaaacagacttaa	1383
Db	481	AGAACATGACCAAGATGTCTTGTTGACTGTGAAGAAGTCCACAGGAAACAAGCTGTGCTAA	540
QY	1384	tcaaggaactcaagccttcagcaaatatcgtgtgacgtgacaggaacacttgaaaaat	1443
Db	541	TCAAGGAACCTCACAGCCTCAGCAAAAGTATATCTGTGATGCGAGAACACCTGANAAGAT	600
RESULT	3		
LOCUS	AL119765	409 bp	mRNA EST 27-SEP-1999
DEFINITION	DKFZp761D0424_r1 761 (synonym: hamy2) Homo sapiens CDNA clone		
ACCESSION	DKFZp761D0424.5		mRNA sequence.
VERSION	AL119765		
KEYWORDS	AL119765.1 GI:5925664		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Bloecher, H., Boeher, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and Wiemann, S.		
JOURNAL	EST (Bloecher, et al.)		
COMMENT	Unpublished (1999)		
	On Jun 22, 1998 this sequence version replaced gi:3247137.		
	Contact: Bloecher H		
	MIPS		
	Am Klopfersplitz 18a D-82152 Martinsried, Germany		
	This is the 5' sequence of the clone insert		
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;		
	sequenced by GBF within the CDNA sequencing consortium of the German Genome Project.		
	No sl sequence available.		
	This clone is available at the RZPD in Berlin.		
	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.		
FEATURES	Location/Qualifiers		
source	1..409		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="DKFZp761D0424"		
	/clone_11b="761 (synonym: hamy2)"		
	/tissue_type="amygdala"		
	/dev_stage="adult"		
	/lab_host="DH10B"		
	/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"		
BASE COUNT	106 a 90 c 114 g 99 t		
ORIGIN			
Query Match	13.4%;	Score 402.6;	DB 64; Length 409;
Best Local Similarity	99.0%;	Pred. No. 8.1e-104;	
Matches 405; Conservative	0;	Mismatches 4;	Indels 0; Gaps 0;

QY	826	ggaggagagagagagagatgtatgatggagagcttcacagacacatgtctgtctgtccaaagat	885
Db	1	ggagagagagagagagagtgatgatgagagcttcacagacacatgtctgtctgtccaaagat	60
QY	886	gggcccgcagcccccacatgatgatgatctctgcgcaaacacacttgcacagagagcttga	945
Db	61	ggggcgcgacgccccacatgatgatgatgtcttcggcaaacacacttgcacagagagcttga	120
QY	946	tgaactagaagattatctatgatgagagcttcttcabatataaagccttatcccatctctac	100
Db	121	tgacctaaagattttctatgatgagagcttcttcacattttaaagccctttaccatttctac	180
QY	1006	cacagtgaaagagagatttaagacagatgtcttcacatttttgatcttcacagccaaagagagac	106
Db	181	cacacgcgaagacgaagatttaccagagctgttctcatatttttgagcttccacgccaaagagagac	240
QY	1066	tgtgttgtttaaaccagaggggaaagaagtacctcctgttacctatctctaaagaagatcagt	1121
Db	241	tgtgttgttttaaccagaggggaaagaaggtacctcctgttacctatctctaaagaagatcagt	300
QY	1126	gaatgtagtcatctttaaagggcaagaggtgtgtgtctgcacccctgcatagtaaagagatgactcgg	1181
Db	301	ggatgtgatgtcatttttaacggcgaagaggggtgtgtgtctgcacccctgcatagtaaagagatgactcgg	360
QY	1186	caagttgaactagtgaaatgatgccccacagagctgtcctcatgtcttta	1234
Db	361	caatgttgaactagtgaaatgatgccccacagagctgtcctcatgtctctttaa	409

Standard Admission At

Accession	Result	LOCUS	Definition
	R59283	446 bp mRNA	EST
	R59283		24-MAY-1995
		y697h01.1	Soares infant brain IN1B Homo sapiens cDNA clone
		IMAGE:41198.5	similar to SP-R2055.5 CE004/9 CAMP DEPENDENT PROTEIN
			KINASE REGULATORY CHAIN + GUANINE NUCLEOTIDE RELEASING FACTORS ;
			mRNA sequence.

ACCESSION	K59283	
VERSION	R59283.1	GI:829978
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
1 (bases 1 to 446)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE	The Washu-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	Contact: Wilson RK
AUTHORS	Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, R., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Scores, M., Tan, F., Trevasaks, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
Insert Size: 451
High quality sequence strops: 315 Source: IMAGE Consortium, LBNL
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
Insert length: 451 Std Error: 0.00
Seq primer: M13Kp1
High quality sequence stop: 315.
Location/Qualifiers
1..446

```

/organism="Homo sapiens"
/db_xref="GDB:414339"
/db_xref="taxon:9606"
/clone="IMAGE:41798"
/clone_1ib="Sceres infant brain INIB
/sex="female"
/dev_stage="73 days post natal"

```

Query Match	Best Local Similarity	Matches 411, Conservative	12.9%, Score 387.2; DB 22; Length 446; 98.6%, Pred. No. 2,1e-99;	0; Mismatches 4; Indels 2; Gaps 2;
OY 2594	ttctatgtcgaggtttgaagaatttaatgtaacccctcaagaacacacagggctcaagctg	2653		
Db 1	TTCTATGGGGAGTTTGAAGTTTATATGACCCCTTCAAGAACCAACAGGGCCTACAGG-TG	59		
OY 2654	acagtagtcaagtcgggaaccctcctctatccctcttaagcttgcctctataaagatatg	2713		
Db 60	ACAGTAGCTAAGCTGGAAACCTCTCTCATGCCCTTTCACAGCCTTTGCATTTAAAGATATG	119		
OY 2714	acatttactcatgagggaaacaagaactcttataacattcaattaaacttgaanaaaatg	2773		

BASE COUNT 142 a 102 c 85 g 116 t 1 others

ORIGIN

Lab host="DB10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lafmld BA; Site.1: Not I; Site.2: Hind III; 1st strand cDNA was primed with a Not I - o1190(dt) primer [5', AACGCGAAGATTCGCGCCGACGAGATTTTTTTTTTTTTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the lafmld BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."

Scorched comparison A

Db 120 ACATTCTACTCATGAGGGGAAACAAGAGCTTCATTGACAACTCTGTAACCTTTGAAAAAATG 179

Db 180 CGCATGTTTCGAATACGCCAGAACAGTGTACTACTACAGAGCCACCTTCATCTCT 239

QY 2834 gatgcagctcaagctataaagaaacatcagatgctccgaggtatgtacggaataat 28939

Db	240	GATGACACTAAGCTAATTAAGAACCTCAGATGTCGGAGTTAATGAGCAATTAAT	299
QY	2894	gtgattacacacagaaacttattacagatgtcacacagattagagcctgtcgaca	2953
Db	300	GATGATTAATACCTACCAATTTATTTATTTTATTTTATTTTATTTTATTTTATTTT	350

2954 tagacattcaaatgcccagaagacagttg -tcctccagtcacacatttccaana 3009
 |||||
 360 TAGACATTTAAATGCCCAAGACAGAGTTTGTCTCCAGTCCACAAATNTTTTCADA 416
 |||||

RESULT	5	1
AU079207		
AU079207	541 bp	
LOCUS	mrna	
	EST	
	20-OCT-1999	

DEFINITION	AU079207 Sugano mouse brain mcb Mus musculus cDNA clone MNCb-3 5', mRNA sequence.
ACCESSION	AU079207
VERSION	AU079207.1
KEYWORDS	GI:6083962
EST.	

SOURCE	ORGANISM
house mouse.	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Rodentia;	
Sciurognathi; Muridae; Murinae; Mus.	
1 (base) 1 to 541	

ADDRESS	INDICED	AUTHORS	TITLE
104000	104000	Hashimoto, K., Kusuda, J., Toyoda, A., Tanuma, R., Ito, A., Hirata, M. Suzuki, Y., Sasaki, M. and Sugano, S.	Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method

JOURNAL
COMMENT
Unpublished (1999)
On Jun 22, 1998 this sequence version replaced gi:3247237.
Contact: Katsuyuki Hashimoto
Division of Genetic Resources
National Institute of Infectious Diseases

23-1, Toyoma 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
Email: khashi@nih.go.jp
URL: <http://www.nih.go.jp/yoken/genbank/>.
Location/Qualifiers


```

source
1. 541
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCB-3986"
/clone_11b="Sugano mouse brain mncb"
/sex="female"
/dev_stage="adult"
/lab_host="TOP10"
/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
was primed with an oligo(dt) primer
[ATGTGGCCCTTTTTTTTTTTTTT] double-stranded cDNA was
ligated to a DraIII adaptor (TGTGGCCCTACTGG) , digested
and cloned into distinct DraIII sites of the pME18S-FL3.
XhoI sites just outside the DraIII sites can be used to
isolate the cDNA insert. Size selection was performed to
exclude fragments <1.5 kb. Library was constructed by
Sugano et al. (University of Tokyo, Institute of Medical
Science). Custom primer for sequencing: 5' end primer
[CTTCTGCTTAAAGCTGCG]"

BASE COUNT      149 a      122 c      125 g      137 t      8 others
ORIGIN

Query Match      11.4%; Score 342.4; DB 69; Length 541;
Best local Similarity 90.8%; Pred. No. 1.4e-86;
Matches 375; Conservative 0; Mismatches 37; Indels 1; Gaps 1.

QY 2390 cagtttgggtgctcactagatcctgcgtttgttctcagctcaagcgttgtacgta 2449
Db 1 CAGTTTTGGGTTTCACCTGANGTCTGCTTTGTTCCAGCTCAGCAACGTTGTCAGACTT 60
|||
QY 2450 ttaaaaaatttataagatagacagcccaactcgttaagagatataaaatcgaattccctt 2509
|||
Db 61 TTGAAAAAATTTTTCAGATGATAGCGGCTCAGTCCAGAGAGATACAAAATTCATAATCCTTT 120
|||
QY 2510 ttggcactgctaclyggactaagtaacattgctgtgagccgcttggcaactaagctggag 2569
|||
Db 121 TTGGCCATCGTCATGGGAGACTCAGCAACGTTGGCCGCTGGAGCCCTTGGCACTTACCTGGGAG 180
|||
QY 2570 aaactggcaagcaagttcaagaagttctatgvgggagtttgaagaatttaatgagacctca 2629
|||
Db 181 AACTGCGCGAGCAAGTTTATAGAGATTCTATGGCGAGTTTGGAGAGCTTGATGATCCTTCC 240
|||
QY 2630 aggaacacagaggcctacagagcgtcagcagtagctaaagcttgaacctcctcctc 2689
|||
Db 241 AGAACAACACAGGGGATACAGGCTGACAGCAGCAGCAAGCTGAGACCCCTTCATCCCTTTC 300
|||
QY 2690 atgcttgcctcattaaagatatagaacttactcgtgggggaacaaagcgttattgac 2749
|||
Db 301 ATGGCCCTGCTTATTAAGATATGACATTTCATGAGGGGAAACAAAGCTTCATTGAC 360
|||
QY 2750 aatctagtaacttgaaaaaatgctcagcagttgcaaatcagcgcagaagacgt 2802
|||
Db 361 AATCTAGTAACCTTTGAAAAAATGGCCATGATTTGC-AACACTGCCAGAACAGT 412
|||

RESULT# 6
AM047433/c 411 bp mRNA EST 18-SEP-1999
LOCUS UT-M-BH1-all-c-07-0-0-UI.s1 NIH BMP4_M.S2 Mus musculus cDNA clone
DEFINITION UT-M-BH1-all-c-07-0-0-UI 3', mRNA sequence.
ACCESSION AM047433
VERSION AM047433.1 GI:5907962
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 411)
AUTHORS Bonaldo,W.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

```

JOURNAL	Genome Res.	6 (9), 791-806 (1996)
MEDLINE	97044477	
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3189695. Contact: Chun, H National Institute of Mental Health 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA Tel: 301 443 1706 Fax: 301 443 9890 Email: MEST@mail.nih.gov	

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized corpus striatum library cDNA library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined.
Seq primer: M13 Forward

POLYA=Yes.

FEATURES
Source location/Qualifiers

1..411
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH1-a11-c-07-0-UI"
 /clone_lib="NIH_BMAP_M.S2"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M.S2 library is a subtracted library derived from NIH_BMAP_M.S1, which in turn is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH_BMAP_M.S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries."
TAG_LIB=NIH_BMAP_M.S2
TAG_RISUB=cornus-stratum
TAG_SEQ=ACGSC"

BASE COUNT	90 a	102 c	90 g	128 t	1 others
ORIGIN					

Query Match	10.8%;	Score 325.8;	DB 64;	Length 411;
Best Local Similarity	88.1%;	Pred. No. 6,6e-82;		
Matches	354;	Conservative	0;	Mismatches 48; Indels 0; Gaps 0.

OY	1954	ccaacctacaacaccattcgggtcgatgcgacttcgtaggaagaagcatcaagtgc	2013
Db	411	cCACACTATATTAACTAATTTCGTGTGGCGTTAGTGCCGCTGGTGAAGAAGTATCATGTGC	352
OY	2014	agttgccgaacaagctcyggtccggaggagcctgatcatagtccaagaataagttccggag	2073
Db	351	AATT	292
OY	2074	agaaaaggctgtcttaaacctaatagtatttgatgatttacgaagtcacacctaatag	2133
Db	291	AGAAAAAGGTGCTCTGAATACTTAATATGTTTGAATTTTACAGCCTCAACATTATATGG	232
OY	2134	acgcctgttcttgcttcgccggagagcaattcgattcagttactcccttaccagaaagga	2193
Db	231	ACGCCCTGTTGGCTCGGCCGAGAGAGCAATTCGACTCACTGACCTCCTTCCCGAACAAGA	172
OY	2194	aggcgaacagtgtggaacagttggaacttttgacagtatgagttccaagaattatgata	2253
Db	171	AGGCCCCGACCACCTGCGCACTGGGAACATTGTTAGCTGATGATGACTCGAAMAGACCTGGCGTA	112


```

/db_xref="taxon:9606"
/clone="IMAGE:1684675"
/clone_lib="Stratagene schizo brain S-11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: Bluescript SK-; Site_1: EcoRI; Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with psychiatric
Diseases (unpublished) Stanley Neurology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."

```

```

BASE COUNT      63 a      68 c      76 g      61 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 268; DB 43; Length 268;
Matches 268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1892 gccacagaagcgcacgctatccgctcgtatgaagttcgttaagtcattgcatg 1951
Db 1 GCCACGAAAGCCGACGCTATCCGGGCTGTATGAATCTGTTAAGGCTATATGCATG 60
QY 1952 gaccacacactacacacacattcgggtccagttgcaccttcgtgaaggaagtcacgt 2011
Db 61 GACACACACCTACACACACATTCGGGTGCCAGTGCCACTTCGGGAGGAAGTCATCAGT 120
QY 2012 gcaattccgacacagctcgtgctcggagggagggcctgatactaatgaagttccgga 2071
Db 121 GCACTTGCCTGCAAGCTGGGCTCCGGGAGGGCTGATCAATGCAAGATGAGTTCGGA 180
QY 2072 ggaagaaagtggtgtcctcaacactaatgattcagatattacgaagctcacattat 2131
Db 181 GGAAGAAAGTGGTGTCTCAACAACTAATGATGATTTCAGATATTACGACCTCACATTAA 240
QY 2132 ggaagcctgttctgtcgtccggagagc 2159
Db 241 GGACGCTGTCTGTCTGCGCGAGAGC 268

```

1626 Score 268 confirmed

```

RESULT 11
LOCUS      A1842094      450 bp      mRNA      EST      14-JUL-1999
DEFINITION UI-M-ANI-afg-e-01-0-UI.s1 NIH_BMAP_MBG_N Mus musculus cDNA clone
ACCESSION A1842094
VERSION    A1842094.1 GI:5476307
KEYWORDS   EST.
SOURCE      house mouse.
ORGANISM   Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 450)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL MEDLINE
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133363.
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
The sequence contained an oligo-dr track that was present in the

```

oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to verify it as a clone from the normalized basal ganglia library (CDNA Library Preparation: M.B. Soares Lab clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. Seq primer: M13 Forward

FEATURES

source

Location/Qualifiers

```

1..450
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-ANI-afg-e-01-0-UI"
/clone_lib="NIH_BMAP_MBG_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73b-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_MBG_N library is a normalized library constructed
from mouse basal ganglia. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dr track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996. Tissue provided by Ms. Annie Novakovich,
Zivic-Miller Laboratories.
TAG_LIB=NIH_BMAP_MBG_N
TAG_TISSUE=basal-ganglia
TAG_SEQ=GTAC"

```

BASE COUNT

```

105 a      107 c      107 g      131 t
ORIGIN
Query Match
Best Local Similarity 93.8%; Score 263.2; DB 61; Length 450;
Matches 285; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

```

```

QY 156 caaactgtagtctg-cctgcagaggtctcttgatgttaagttatcagaccagcagtc 214
Db 450 CCAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 391
QY 215 accagagctgttgaccatctgaccctgggaattgggaagccttggaagttccatc 274
Db 390 ACCAGAGTCCGCTGACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 331
QY 275 tgacacacaccccgccatgcaacatcgttacccagagagagcagtgactgtccgca 334
Db 330 TGGATTAACACCCCTGCGCATGCAACATCGTTACCAAGGAGAGCGCAACTTCTCCGCA 271
QY 335 tcgagcagaagacttcaagcactatggagaaatcagcagatatactgacagattc 394
Db 270 TTGAGCAGAGGAGACTTCAAGGCACTATGAGGAATAATACCAAGATATATGCGGACTTC 211
QY 395 tggctcctctatgttatatggaaagggctcaacaaatgacagaggttccctcaagg 454
Db 210 TGGCTCTCTCCATGATGTATGGAAGGGCTCTTAACAAATGACAGGATTCCTGCAAGG 151
QY 455 aaga 458
Db 150 AGAA 147

```

```

RESULT 12
LOCUS      A1608196      277 bp      mRNA      EST      21-APR-1999
DEFINITION vc85b01.y1 KO mouse embryo 11 5dpc Mus musculus cDNA clone
IMAGE:789769 5' similar to IR:Q92565 Q92365 MTEOBLAST KIAA0277. ;,
mRNA sequence.
ACCESSION A1608196
VERSION    A1608196.1 GI:4617363

```

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 277)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The Mashu-NCI Mouse EST Project 1999
Unpublished (1999)
On Jun 22, 1998 this sequence version replaced gi:3247157.
Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

FEATURES
source This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:474177
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
Seq primer: Primer name ambiguous
High quality sequence stop: 268.
Location/Qualifiers
1..277
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:789769"
/clone_lib="Ko mouse embryo 11 5dpc"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="11.5dpc"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Total RNAs were extracted from 11.5 dpc embryos (excluding placenta and yolk sac). The double-stranded cDNA was synthesized with an oligo (dT)-1 primer GAGAGAGACCTGTTCTAGATGCCAGCGCGCTTTTCTTTTCTTTT 3'. The cDNAs were ligated to L1-Sal3A: 5' GCTATGACGTCGACTATCC 3' and L1-Sal3B: 5' GATATGCGAGCTCAAT 3'. The cDNAs were size-selected and amplified by long-range PCR using Ex Taq polymerase for 18 cycles. The PCR-amplifiable cDNA mixture went through one round of equalization and was digested with SalI/NotI and cloned into the SalI/NotI sites of the pSPORT1 plasmid vector (Life Technologies). The library was constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong Wang."

BASE COUNT 83 a 72 c 59 g 63 t
ORIGIN

Query Match 8.0%; Score 239.8; DB 48; Length 277;
Best Local Similarity 92.0%; Pred. No. 1.9e-57;
Matches 253; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2595 tctatcggaagttgaagttaatgagcccttaagaagaacagagcctcaagctga 2654
|||||
DB 3 tctatcggaagttgaagttaatgagcccttaagaagaacagagcctcaagctga 62
|||||

QY 2655 cagtaagctgaagcctcctcatccctcctccttcgttccttcattaaagataa 2714
|||||
DB 63 CAGCAGCCAGAGCTGAGAGCCCTCTCATCCCTTTTCATGCCCTTCTTTTAAAGATATA 122
|||||

QY 2715 cattactcatgaggggaacagctcattgacaatctagtaaaacttgaaaaatgc 2774
|||||
DB 123 catttactcatgaggggaacagctcattgacaatctagtaaaacttgaaaaatgc 182
|||||

QY 2775 gcatgattgcaatagcgccagagaactgagatactacagagcaacccctcatccg 2834
|||||
DB 183 GCATGATTGCAAAACACTGCCAGAACACTACGGTACTACAGAGCCAGCCCTTAAATCCGG 242
|||||

QY 2835 atgcagctcaagctaatataagacatcagatgctc 2869
|||||
DB 243 ATGCGCGCTCAAGCTATATAGAACCATCAGCATGTC 277
|||||

RESULT 13
AI386336
AI386336
LOCUS
DEFINITION mRNA
AI386336 441 bp
AI386336
VERSION
KEYWORDS
SOURCE EST.
ORGANISM house mouse.
MUS musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 441)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The Mashu-NCI Mouse EST Project 1999
Unpublished (1999)
On Jun 15, 1998 this sequence version replaced gi:3225813.
Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

FEATURES
source This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:332812
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Glibco
High quality sequence stop: 376.
Location/Qualifiers
1..441
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:552020"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. 6-8 month old female lung and 1.5 year old male lung were source of mRNA. Average insert size: 1.5 kb; UniZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT 132 a 112 c 90 g 106 t 1 others
ORIGIN

Query Match 7.3%; Score 220.4; DB 45; Length 441;
Best Local Similarity 68.8%; Pred. No. 7.8e-52;
Matches 302; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 2388 ttcaagtttggtcgctcaactgagatcgtctgtctcaagcagcaagcgtgtcagc 2447
|||||

Contact: Simpson A.V.B.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPSP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/seq/gethtml.pl?ref=IL-BR202-011.html&ct3=2503996&ct4=1>)
 Seq primer: puc 18 forward.
 Location/Qualifiers
 1..540
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="BR202"
 /sex="female"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18, Site_1: Sma1; Site_2
 Sma1; A mini-library was made by cloning products derived
 from ORSITES PCR (U.S. Letters Patent application No.
 196,716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue

04-290 IN CIRCULAR NOT FOUND, NOT A SITE SHOWN IN BEGINNING OF SEQUENCE

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 17:24:28 ; Search time 13270 Seconds

(without alignments)
-689.422 Million cell updates/sec

Title: US-09-422-999-17

Perfect score: 3013
Sequence: 1 gatccagcagatgtgtgat.....ccacaatttcaaaatgcc 3013

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl3:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vl:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vl:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pt4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*

45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2957.4	98.2	4278	40 HSU78516	U78516 Homo sapien
2	2413.6	80.1	2982	12 AF115480	AF115480 Mus muscu
3	1112.8	36.9	2399	12 RN078517	U78517 Rattus norv
4	721.6	23.9	3394	40 HSU78168	U78168 Homo sapien
5	720	23.9	2646	40 AF103905	AF103905 Homo sapi
6	636.6	21.1	3373	12 RN078167	U78167 Rattus norv
7	393.2	13.1	5900	9 D87467	D87467 Human mRNA
8	176.2	5.8	59255	44 AC016789	AC016789 Homo sapi
9	141.8	4.7	87767	43 AC014497	AC014497 Drosophill
10	141.8	4.7	111455	33 AC007549	AC007549 Drosophill
11	141.8	4.7	189668	33 AC007624	AC007624 Drosophill
12	132.4	4.4	125681	41 AC008340	AC008340 Drosophill
13	117.6	3.9	59255	44 AC016789	AC016789 Homo sapi
14	102.6	3.4	209071	35 AC005285	AC005285 Drosophill
15	99.2	3.3	6568	9 AB002311	AB002311 Human mRN
16	94	3.1	48096	34 CER2065	Z30423 Caenorhabdi
17	92.6	3.1	128473	11 AC004241	AC004241 Homo sapi
18	81.2	2.7	152184	32 CER70610	AL020987 Caenorhab
19	71.4	2.4	110000	33 AC005079_3	Continuation (4 of
20	71.4	2.4	224064	33 AC005055	AC005055 Homo sapi
21	64.2	2.1	170128	33 AC005059	AC005059 Homo sapi
22	63.6	2.1	111455	33 AC007549	AC007549 Drosophill
23	54.2	1.8	22920	34 CER14610	Z68880 Caenorhabdi
24	52.8	1.8	85181	11 AC004622	AC004622 Homo sapi
25	52.2	1.7	2035	34 DD053884	U53884 Dictyostell
26	50.6	1.7	5398	7 YSCDC25	M15458 Yeast (S.c
27	50.6	1.7	5548	7 SCCDC25G	X03579 Yeast cell
28	50.6	1.7	19198	8 YSC12142	U17247 Saccharomyc
29	50.6	1.7	125681	41 AC008340	AC008340 Drosophill
30	46.2	1.5	3087	12 RNO249986	AJ249986 Rattus no
31	45.4	1.5	2919	8 AF024633	AF024633 Magnaport
32	45.4	1.5	3172	8 AF015753	AF015753 Magnaport
33	44.8	1.5	7218	5 I66494	I66494 Sequence 14
34	44.4	1.5	4278	40 HSU78516	U78516 Homo sapien
35	43.6	1.4	4062	5 E08769	E08769 cDNA encodi
36	43.6	1.4	4070	9 H0MC36P	D12139 Homo sapien
37	43.2	1.4	3644	7 YSCDC25A	M82964 Saccharomyc
38	43	1.4	224064	33 AC005055	AC005055 Homo sapi
39	41.8	1.4	3731	34 ANDDC13	D28663 Sea urchin
40	41.4	1.4	3711	40 AF023130	AF023130 Homo sapi
41	41.4	1.4	3711	40 AF181250	AF181250 Homo sapi
42	40.8	1.4	163415	41 AC010756	AC010756 Homo sapi
43	39.4	1.3	3268	7 SPST6	X53254 S.pombe ste
44	39.4	1.3	29500	7 SPC1442	AL031966 S.pombe c
45	39.4	1.3	47385	7 SPC1450	AL049559 S.pombe c

ALIGNMENTS

RESULT 1
HSU78516 HSU78516 4278 bp mRNA
LOCUS Homo sapien CAMP-regulated guanine nucleotide exchange factor II
DEFINITION (CAMP-GEFII) mRNA, complete cds.
ACCESSION U78516
VERSION U78516.1 GI:4115908

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiinae; Homo.
1 (bases 1 to 4278)
Kawasaki, H., Springett, G. M., Mochizuki, N., Toki, S., Nakaya, M.,
Matsuda, M., Housman, D. E. and Graybiel, A. M.
A family of cAMP-binding proteins that directly activate rap1
Science 282 (5397), 2275-2279 (1998)
99074384
2 (bases 1 to 4278)
Springett, G. M., Kawasaki, H., Graybiel, A. M. and Housman, D. E.
Direct Submission
Submitted (18-NOV-1996) Center for Cancer Research, Massachusetts
Institute of Technology, 77 Massachusetts Avenue, E17-540,
Cambridge, MA 02135, USA
On Jan 8, 1999 this sequence version replaced gi:4106959.
Location/Qualifiers
1..4278
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q31-q32"
1..4278
/gene="cAMP-GEFII"
109..3144
/gene="cAMP-GEFII"
/codon_start=1
/product="cAMP-regulated guanine nucleotide exchange
factor II"
factor II"
/protein_id="A003422.1"
/db_xref="GI:4115909"
/translation="MVAHAHSSSSAEMIACTDKRRLRESEEDVLIIFRLKEVAF
EKHPNLIHQICGYENLEKGIPLRGDIDGTMNAYVLASLDVKKVETSSHOAY
TICTLIGTARFSGSIIDNTPRHATITRRESEILTRQEDFKALMKRYQTMAGLIAP
PIGVMEGSSNDRIPIKENTPLIEPHVPLRPATITIKVSEKILRAGKILRNALISRA
PHMIRDRKYLKTYRQCCVGTGLVDMOQTPCHSRTOAVGMMOVLTDGVLNHDQ
ENHFOKYLFRFLDEHEDAPLPTEEKKECEDELDMLLSQGPRAHMLIRK
PGQRTVDLEIIEYELHLKALSHLSTYKRLAVLIFESAKGTVLPFNGEGCT
SWYILKGSYNNVYIGKGVCTLDEGDDGKLAVIDAPRAASTYVRENGCFPIYDK
EDPNRLIRVEANTVRLKEDDOVLYLEKYPAGNRASNGNSOPKQITVMSGTPKTI
LEHLETLRLAATLNTDSVLDFTIMHCVEPNQQLCPALVAHHAPOSGTEDEK
MDYALNNRRVIRLVQMAAMGDLQEDVDNAFLDEEYVSDERLAIALKEQLP
ELEKIVQISEDAKQKHKVLLQOFNTGDEARQROPIRGSDVLFVYCMNDHYT
TIRVPATSYKEVISAADKISGEGELIIVKMSGGEKVLKPNDSVPTTLINRL
FACRPOFOSIPLPEOGEPTVGTGTFELMSKDLAYOMTYDMLPCYHDELIIY
HTGRNREKTTANLDELRLRNEIOPWVYTEICLSQSLKRYQLKRTIAHCKE
YKUNSEFALVGLSNIAVSRLLATWEKLPKSKRYAEFESLMDSRNHRVRLTVA
KLEPPLIPEMPLIKDMITFHEGNKTFIDNLVNFEMRMIAANTARTVRYRSQPFMPD
AAQANKHNOVRSYRQNLNVIDNORTLSQMSHRLERPRD"

BASE COUNT 1273 a 903 c 963 g 1139 t
ORIGIN

Query Match 98.2%; Score 2957.4; DB 40; Length 4278;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2998; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

Db 365 GGTCTTGGATGTTAAAGTATCTGAGACCAGCAGTCCAGGATGCTGTGACATCTGTA 424
Qy 239 CCTGGGAATTGGAGCGCTTTGGAGATCCATCTGTGACAAACACCCCGCATGCA 298
Db 425 CCTGGGAATTGGAGCGCTTTGGAGATCCATCTGTGACAAACACCCCGCATGCA 484
Qy 299 ccatcttaccaggagagcaatgcaatgctccgcacatcagagagagagcttcaaggc 358
Db 485 CCATCTTACACAGAGAGACAGTCACTCTCCGATTCAGCAGAGAGACTTCAAGCCAC 544
Qy 359 tatcgagaagaatcagacagatataatgacagagcttcctccctatgattatg 418
Db 545 TATGGAGAGAAATATCCAGATATATGAGAGACTTCTGCTCTCTCTATGCTTATAG 604
Qy 419 aaacggccttaacaatgacagatctccgcagaagaagacacac--ctcatgacatc 476
Db 605 AAACGGCTCTAACAATGACAGAGATTCCGACAAAGAGAACACCTCTCATATTGAACCTC 664
Qy 477 acgtctctctgctcgttaacacacatatacgaagctccctcagagagagatccacag 536
Db 665 AGCTTCTCTGCTGCTCTGCTTAACACATTTACAGAGTCCCTTCAGAGAAATCCTCAGAG 724
Qy 537 ctggaanaatttaacgaatgcattctctctcgagcactacatgataagataga 596
Db 725 CTGGAANAATTTTACGAATGCCATTTCTCTGAGCACCCTCATGATTAAGATAGAA 784
Qy 597 aatcacacctaaagacatacagaacatgctgtgtggaactaaactgtgtgactgataga 656
Db 785 AATCACCACTTAAGACATACAGACAAATGCTGTGTGGAACTAACTGGTGGCTGATGA 844
Qy 657 tcgacgagacacacatgcttcaactcccgagctcaagctgtgtgcatgtgcaagctct 716
Db 845 TCGACGAGACACCATGTTGTCATCTCCGAGCTCAAGCTTGTGGATGTGGACAGTCGT 904
Qy 717 tagaagaatgcttctcaacacgctgagccaggagacacattccaagaacttattat 776
Db 905 TGAAGATGTTGTTCAACCAACGATGAGACAGACACCATTTCCAAACAAATATTAT 964
Qy 777 tcatcgaattctgagatgagcagagagatgcccttgcctactaagaaggagaaga 836
Db 965 TCTATGATTTCTGATGATGACACGAGATGCCCTTGTGCTGCTGAGGAGGAAGA 1024
Qy 837 aggaatgataagagagctccagagacacacatgctgtgtgtcaacagatggccccag 896
Db 1025 AGGAGTGTATGAGAGCTCCAGAGACACATCTCTGTGTCAAGATGGCCCCAGC 1084
Qy 897 cccacatgagatgatactctcgcaacacacctgagcagagagctgtgatacctaga 956
Db 1085 CCCACATGAGATGATCTTCCAAACACCTGGCCAGAGAGCTGTGATGACCTAGAGA 1144
Qy 957 ttatcctagaagcttctcatattaaagcttatccatcttctccacaggaac 1016
Db 1145 TTATCTATGAGAGCTTCTTCATATTAAAGCTTATCCATCTTCTACCAAGTGAAC 1204
Qy 1017 gagaatgacagatgcttctcatatttgaatctcaagccaaaggagagagctgtgtt 1076
Db 1205 GAGAGTATCAGAGTGTCTCATTTTGAATCTCACAGCCAAAGAGAGAGCTGTGTTA 1264
Qy 1077 accaaggaggaagagatctctcctgtgatacatattctaaagatacagtgaaatga 1136
Db 1265 ACCAGGGGGAAGAGATCTCTCTGATCTTCTTAAAGATCACTGATGATGATCA 1324
Qy 1137 tttaaggcaagaggtgtgctgcacccctgcatgaagagagatgacttggaattagaac 1196
Db 1325 TTTACGGCAAGGGTGTGCTGCTCACCTTGATGAGAGAGATGACTTGGCAAGTTAGAC 1384
Qy 1197 tagtgaatgatacccaagagctgctctatgcttcaagaaagataactgcaattct 1256
Db 1385 TAGTGAATGATCCCCACAGAGCTGCTGTATGCTTACGAGAAGATTAACGTGCAATTC 1444
Qy 1257 taagatgagacaagagatctcaacggatctcaaggaacgttggaggcgatacagta 1316
Db 1445 TAAAGATGACAAAGAGATTTTCAACCGATCCTTAAGGAGAGCTGAGCGAATACAGTCA 1504

OY	1317	gaactaaagaacatgacccaagatgctctggctcgtgagagaaggtcccgacagyggaacagag	1376
Db	1505	GACTTTAAAGAACATGACCAAGATGCTTGCTGGTCCGAGAAAGGTCCCAACAGGAAACAG	1564
OY	1377	cttctaacaagaaactcaacgcccacgacaagaatacagctgtgacagtaacacgctg	1436
Db	1565	CTTCTATCAAGGAACCTCACAGCCCTCAGCAAAAGTATCTGTATGTCCAGGAACACTCG	1624
OY	1437	aaaaaatttagagcaattctagaacaacaacgccttgaaggaacatttaagagaa	1496
Db	1625	AAAAAATTTTAGAGCACTTTTCTGAAACAAATACGCCCTTGAGGCACTTTAATGAAGCAA	1684
OY	1497	cagatctcgttttaaatgaactttatataatgaatgaacgtggtttttagccaatccccagc	1556
Db	1685	CAGATCTCGTTTAAATGAAGCACTTTATATATGACACTGTGTATTATATGCAAAATACCAAGC	1744
OY	1557	cttcaccggaactggtggtgcccactaccagcgaagcgccttcaacaagtgtaacagaacagaga	1616
Db	1745	TTTGCCCGGACGTGGTGCCCACTTACCAGCAAGCGCTTCACAAGGTACAGAACAGAGGA	1804
OY	1617	aaatggaattatgcccctcaacaataaagaagcgagatcatccgcctgtttctcaagtygctg	1676
Db	1805	AAATGGATTATGCCCCCTCAACAATTAAGCGCGATCATCCGCTGTCTTACAGTGGGCTG	1864
OY	1677	ccatgatatgagaacctccctcacaagagatgaacatctatctgtgcctccctcggagagattt	1736
Db	1865	CCATGTTGAGGAACCTCCTCTCAGAGAGATGACGTATCTATGGCCCTTCTCGAGGAGTTTT	1924
OY	1737	atgatctgtatcacagatgatgcgcgagatgatgctgcctcctaagaagaaactgcccagagt	1796
Db	1925	ATGATCTGTATCGATGATGATGCCCGGATATGTCCTCCCTCAAGGAGCAACTGCCCAGT	1984
OY	1797	tggagaaagattgtcaagcaaatctcgaagaatgacaaggcaccaaaaagaagacaagag	1856
Db	1985	TGGGAAGATTGTCAACCAATATCTCGAAGATGTCAAAAGGCAACCAAAAGAGCAAAAG	2044
OY	1857	tctctttgcaacaagttcaacacggcgagatgagagagcccaagaagcgcaagctatccgcg	1916
Db	2045	TTCTTTTGCACAACTTCAATATACGGGGGATGAGAGACCCAGAAAGGCCAGCCTATCCGCG	2104
OY	1917	gctctgatgaagttcgttctaaggtctatctgatgacacaactacaacaacattcggg	1976
Db	2105	GCTCTGATGAAAGTCTCTGTTTAAGGTCTATTTGATGACCAACCTACCAACCAATTCGGG	2164
OY	1977	tgcacgtaggcgaactcggtagaagaagatcaacagtcagttgcgacaagctcgggctcg	2036
Db	2165	TGCCAGTGGCCACTTCCGCTAAGGAAGATCTCATGCTCAAGTTCGCAAGACTGGGCTCCG	2224
OY	2037	ggagagggcctgatcatcatagtaagaatgattccggagagagaaaagttggtcctaacaacta	2096
Db	2225	GGGAGGGGCTCATATAGTCAAGATGAGTCCGAGAGAAAAGGTGGTCAAACTCA	2284
OY	2097	atgatggttcaagatatttaagaagctcaacataatgaagagcgtgtgttgcgttgcgcgag	2156
Db	2285	ATGATGTTTCAGTATTTAACAACGCTCAACCTATATATGAGACCGCTGTTCGCCGAG	2344
OY	2157	agcaatctgatcaactcgaactcccttaccagaagaaaggaagcgccaactgcttggaacaagtg	2216
Db	2345	AGCAATTGCAATTCAGTCACTCCCTTACCAAAACAGGAAGGCCCACTGTTGGAACAGTGG	2404
OY	2217	gaactttgaactgaatgaagctccaagaatttagcaatacagaatgacaaattatgtatgg	2276
Db	2405	GAACTTTGAACTGATAGCTCCAAATAATTTACATACAGATGACAAATTTTGATTGGG	2464
OY	2277	aaacttcaactgctgacagagcttgtaggtaattcatcacacaatttgaaggaactaat	2336
Db	2465	AACCTTTCAACTGGCTGCAATGAGCTGAGCTAATCTATCATCAACATTTGGAAGCAATATT	2524
OY	2337	ttaaaagaaccagcaacttgattgttctcctgagagatttaagaaattcaattcaagttt	2396
Db	2525	TTAAAAACACCAAGCAAACTTGATTTGTTCTCGAGGAGATTTATATGAATTCAGTTTT	2584

QY	2397	gggctgcacatgagatctgcctctgttcttcagctcagcaagcgtgttcagctatataaa	2456
Db	2585	GGGTGCTACTGAGATCTGCCTTTGTTCTTCAGCTCAGCAAGCGTGTTCAGGATTATTA	2644
QY	2457	aatttattaagaatcacagccaccctctaaaggagatataaaatctaatccctttttgcca	2516
Db	2645	AATTTATTTAGATACAGCCGCACTTAAGAGATATTAATAATCTGATATTCCTTTTGGCA	2704
QY	2517	tgcgtacggacataagatacatctgtctgtgagccgcttgcacatacgttgggagaaactgc	2576
.Db	2705	TGCTCATGGGAGATTAGTAAACATTTGCTGTGAGCGCGTTTGCACTAACGTTGGGAAACTGC	2764
QY	2577	caagcaagttcaagaagttctatcgcgagtttgaagtttaatggaacccttcaagaaacc	2636
Db	2765	CAAGCAAGTTCAAGAAAGTTCTATGCGGAGTTTGAAAGTTTAAATGACCCCTTCAAGGAAC	2824
QY	2637	aaaggagccacaggtctgacagtagtaagatgcygaaccctccctctatccctctaatgcctt	2696
Db	2825	ACAGGGCCCTACAGGCTGACAGTAAAGCTGAAGCTGGAACCTCTCTCATCCCTTCATGCTT	2884
QY	2697	tgcctacataaagatatgacatttacttcatatgaagggaacaagaagctttcatatgaacatcag	2756
Db	2885	TGCTCATTTAAATATATGACATTTACTCATGAGGGGAACAAGACGTTTATGACATTTAG	2944
QY	2757	taaaccttgaaaaaatgcygatgattgcaaatatagcgccaagacagttgagatactacagga	2816
Db	2945	TAACTTTGAAAAAATGCGCATGATTTGCAAAATPACGGCCAGAACAGTGAATACGTACAGGA	3004
QY	2817	gccaacaccttaaatcctgtatgtcagcttcaagctcaataaagaacatcaagatgtccggaatt	2876
Db	3005	GCCAAACCTTCATCCTGTATGTCAGCTCAACGCTAAATPACGAACCATCAGATGTCGGAATT	3064
QY	2877	atcgtacggacataaataatgcatgtgacgaacacagagaaccttatcacagatgtacacagat	2936
Db	3065	ATGTACGGCAATTAATATGATTTGATGACAAACACAGAACTTTATACAGATGTACACAAT	3124
QY	2937	tagagcctcgtcagacatagacatttccaatgccccaaagcaacagtttgtctccagttcca	2996
Db	3125	TAGAGCCTCGTGCACCATATGACATTTCAAAATGCCAAAGCAACAGTTGTCTCCAGTCCA	3184
QY	2997	caatttcaaaaa 3009	
Db	3185	CAATCTTTCAAAA 3197	
RESULT 2			
AF115480	2982 bp	mRNA	ROD 25-JAN-1999
LOCUS	Mus musculus CAM- dependent Rap1 guanine-nucleotide exchange factor		
DEFINITION	mRNA, complete cds.		
ACCESSION	AF115480		
VERSION	AF115480.1	GI:4185566	
KEYWORDS	house mouse.		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
REFERENCE	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2982)		
TITLE	Gaudriault,G.E., Takaya,K. and Vale,W.W.		
JOURNAL	A brain cAMP-dependent Rap1 guanine-nucleotide exchange factor		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 2982)		
TITLE	Gaudriault,G.E., Takaya,K. and Vale,W.W.		
JOURNAL	Direct Submission		
AUTHORS	Submitted (19-DEC-1998) The Clayton Foundation for Peptide Biology,		
TITLE	The Salk Institute, 10010 N. Torrey Pines Road, La Jolla, CA 92037,		
JOURNAL	USA		
FEATURES	Location/Qualifiers		
source	1..2982		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/tissue_type="Brain"		
CDS	1..2982		

				/codon_start=1 /product="cAMP-dependent Rap1 guanine-nucleotide exchange factor" /protein_id="AAD09132.1" /db_xref="GI:418567" /translation="MVAAHAHSOSSAEMWIAICLDRPLERSSEVDVLIITRLKRAVEKFPNLRIOICLCGYENLEKGTITLERODIGTNNVAVLASLDKVSSTSHODAP TICTLIGTAFGEISLDNTPRHATIVYRESELRLBOEDFKALMEKRYOYMGILAP PYGMETGSSNDRIPDKENYPSKEITIRAGKITLILALISRA PHILRDKYHLKTYRCC VGTPLVPMHIOQTSCHVSRIOAVGMOMVYLEDEVLYNVDERHFQDYLYEPLRDLDER EDAFLPTBEERKECDEELQDTMLLSQMGDAIMRLIRPQGRVDDLEIYYIDEL HIKALSHLSTTVRKELAGVLFESHAGGCTVLFNQGEGSWYIILKGSVNVITYGK VVCTLHGEDDFEKLALVNDAPRAASIVLRDNQELRLDEKEDNRIILDEAATVBLK EHODVLYLEKVPNRANAGONSPOOKRTWMSGPEKILHEHLETRIPESTINAT DSVLNDPMVHCVMPNTOICPALVIAHYHQPSQGTPEOEMDALNKRKRVILVLIOW AAMTGDILQEDDVAMAFLEEFYVSDDAAMAFKEQLEPEKIVKQISEDKAPQK KHHVLIQGFNTGDERAKRQPIKGSDEVLKVCIDHTYTTIIVPAASKEVYSVA DKLSGEGLIIVKMSNGEKVVLKSNDSVFYTLTLTNFLAFCRPOFSLPPLPOE GPTTGIVGTELEMSKDLAYOMTTYDMELEFNCHGHELIYHTFGRNPKKTKTNDLF LBRNEIQEFWVTVEVCLQSOLSKRVOLIKFKITIAHCKEYKMLNSFPALVMSLVA VSRILATMEKLPKFKFYAEFSLMDPSNHAIRLTAKEPLIPEPDLIKDMT FTHBCKTEIDNLFNEKMRMIANTATVTRYSPNPDAQANKNHODVRSYROL NVIDNQTSIQMSHRLEPRP"	
BASE COUNT	824 a	685 c	758 g	715 t	
ORIGIN					
Query Match	80.18;	Score 2413.6;	DB 12;	Length 2982;	
Best Local Similarity	89.68;	Pred. No. 0;			
Matches 2650;	Conservative 0;	Mismatches 254;	Indels 54;	Gaps 3;	
QY	1	gacacagcaagatcgttgatataatcttcactcagcaagaagttaaagcttttga	60		
DB	77	GATCAGAGGAGATGTGGACATAATTTTACGCGCGCTGAAGGAGTTAAAGCTTTGAGA	136		
QY	61	aattccaccaatctcccttcatacagatttgctatgttgattatagaaatcttgaaa	120		
DB	137	AATTCACCCAAACCTCCTTCGTGAGATTGTGTATGCGGTTACTATGAGACCTGGAAA	196		
QY	121	aggaataacatttttcgcgaaggtgata-tgaaacaactgtgatctg-cctggcag	178		
DB	197	AAGGAATCACACTGTGTGCGCAAGGAGATATGGAACCACTGGTATGCTGTGCTG	256		
QY	179	gctcttgatcgttaaaagtatctgagaccagcagctcaagatgctgtgacatcgtga	238		
DB	257	GGTCTTGGATGTTAAAGTGTCTAGACACGACAGTCAACGAGATGCGGTAAACATCTGCA	316		
QY	239	cccttggaatttggacggtcctttggagagtcacatctgcaacaacaccccgcatgca	298		
DB	317	CTCTGGGAATTTGGGACAGCCTTTGGAGAGTCGATTTGATTAACACCCCTCGCATGCAA	376		
QY	299	ccatcgtttaccaggaagagcgtgaactgctcgcatcgacgcaagaagacttcaaggac	358		
DB	377	CCATCGTTACCGAGGAGAGCGAACCTTCCGCATTGAGCAGAGAGACTTTAAAGGCAC	436		
QY	359	tatggagaaatcgcagcagtatatgacagacttcgctccttccttatggtatata	418		
DB	437	TATGGAGAAATACCGACAGTATATGGCCGACCTTCTGCTCTTATGGGTATATGG	496		
QY	419	aaacggcctcaacaatgacaagattcctgacaagaagaaacacatcattgacatcac	478		
DB	497	AAAGGGGCTCTAACATGACAGGATTCCTGCAAGAGAAAT-----	538		
QY	479	gttcctcttcgtcctgtaacaacatcaagaagtccttccttcagagaagatcctagact	538		
DB	538	-----GTCCCTTCAGAGAGAGATCCCTCAGAGCT	564		
QY	539	ggaanaatttacgaaatgcattctctcgcagcaccctacatgtaagaagataaaaa	598		
DB	565	GGAAAAATTTTACGAATGGCATTTCTCTCGAGCTCCCCACATGATAAAGACACAAAG	624		
QY	599	taccacctaagaacatacagacaatctgtgtggaactgaactggctgacatgcgac	658		
DB					
DB	625	TACCACCTTAAGACATACAGACATATGCTGTGTGGACTGAGCTGGTAACTGGATGATA	684		
QY	659	gacagacacacatgtgttcacatcccgactcaagcgtgttgacatgtgcaagtcctgtta	718		
DB	685	CACACAGACATCCTGTGTTCACCTCGCGAGCTCAAGCTGTGGCATGGCAAGCTTGCTG	744		
QY	719	gaagatggtgttcttaaccaacgtgagccgggagacacatttccaagacttattatc	778		
DB	745	GAAGATGgtgttcttccAACCATGTGGACGAGACGGCCTTTCMAAGCAAAATATTTATTT	804		
QY	779	tatcgaatttcgtatgatcagcagcagagatgcccccttgcctactagtaggagagaag	838		
DB	805	TATCGATTTCTGGATATACAGCTGTGAGATGCCCTTTGCTTACTGAGSAGAAAGAAAG	864		
QY	839	gagtgatgtaggaagctccaagacacatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	898		
DB	865	GAGTGTGATGAAGAACTTCAGAGACACCATATGCTGCTCTCACAGATGGGCTTGACGC	924		
QY	899	caatgagagatgactccttcgcaaaccaactgagcagagagatcgtgtgataagat	958		
DB	925	CACATGAGATGATCCTGGAAAACCATGGCCACAGAGACTGTGATGACCTAGAGATT	984		
QY	959	atcatgagagactcttcataatataaagcctatcccatcttcacacagtgaaacga	1018		
DB	985	ATCTAGACAGAGCTCCTTATATTAAGCCTTATCCCATCTCTTACACAGTGAACG	1044		
QY	1019	gaagttagcaggtgttctcatatttgatgtcagcccaagagagagagaggtgtgtttaa	1078		
DB	1045	GAGTTAGCAGGTGTCTTCATTTTGTAGTCTCACGCCAAGGAGAACTGTGTGTTAAC	1104		
QY	1079	cagggggaagaggtacccctcctgtgtacatattcttaaaagatcagtgatgacat	1138		
DB	1105	CAGGGGGAAGAGTACCTCTGTGTACATCATCATCTGAAGAGATCCGTGATGTACTAT	1164		
QY	1139	tacgcaagaggtgtgtgtcgcacccctgcataagaagagatgacttcgcaagtgaccta	1198		
DB	1165	TATGGCAAGGTGTGTGTCTGCAACCTGCAACGAGAGAGATGATTGGCAAGTACTCTA	1224		
QY	1199	gtgaatgtagcccccagagctgcctctacatctcgtcttaagagaataactgccaattcta	1258		
DB	1225	GTGMAAGATGCTCCCAAGAGCTGCTCCATATTTCTTCGGAGAGATTAATGTCACTTCCA	1284		
QY	1259	aaagatgcaagagagatctcaacggatcctcaaggagcgttgaaagcgaataacgtcga	1318		
DB	1285	AGAGTGCACAAAGGAACCTTCAATCGGATCTGAGGAGCTTGAGGCCAATACATCTGCA	1344		
QY	1319	cttaagaacatgacaagaatgtctgtgtcgtggaagagctccagcagggagagagct	1378		
DB	1345	CTTAAAGAACATGACCAAGATGTCTTGTGACTGAGAAAGTCCGACGAGGAGACGACT	1404		
QY	1379	tctaatcaaggaactcacagcctcagcaaaagtatactgtgtgtgtgtgtgtgtgtgtgt	1438		
DB	1405	GCTAAACAAAGAACTCAACAGCCTCAGCAAAAGTATACGTGATGTACAGAAACACTGAA	1464		
QY	1439	aaaatttgaacatcttcttaaaaaaataagcctcttgaggcgaacttaatgaagaaca	1498		
DB	1465	AAATTTTGAACCATTTCTTGAAGCAATACGCTTTGAGCCATCTTTAAATGAACCAACA	1524		
QY	1499	gattcgttttaaatgacttataatgtacatgtgttttataatgcaataaccagatt	1558		
DB	1525	GATTCGGTTTAAATGACTTGTATGATGACATGTGTTTATATCCAAATATCCAGCTT	1584		
QY	1559	tgccggagactgtgtgccaactacacgcaacagccttcacaagtgatcacagaaagagaa	1618		
DB	1585	TGCCCTGCTTGTGGCCATATACACGACACAGCTTCTCAAGGATACGACAGAGAGGA	1644		
QY	1619	atggaatttagccctaaataaagagcagatcatccgctcgtgttctcaatgagctgcc	1678		
DB	1645	ATGGATTTATGCCCTCAACAAACAAAGAGGCGGTATCCCTCTTGCAATGGGCGGCC	1704		
QY	1679	atgtatgagacctcctgcaagaagatgactatcatgagccttcctggaagagtttat	1738		
DB	1705	ATGTATGGCGATCTCTCCAAAGAAAGATGTGGCATATGGCTTCTTGAGAGATTCTAT	1764		

OY	1739	gtaacgtatcacgaatgatcccgagatgatgtctgcctccataagagacaatgtgccaagtgtg	1798
Db	1765	gtgtctctatcaaatgatgcacgcagatgatgtgcctcttcaagagacagctccagacgtg	1824
OY	1799	gagaaagtctcaagcaaatctcaagaagaatgtcaaaagacacacaaagaacacaaagtt	1858
Db	1825	gagaaagtgttcanaagcaaatctcagaaagcgaanaagctccacagaaacacaaagtg	1884
OY	1859	cttttgaacagttcataacacgggcgaatgaagagcccaagacgcacagccatccggcg	1918
Db	1885	ctttttgcaacagttcaaacacagagtgacgagggcccgaaagccctacgcttattcgtaggc	1944
OY	1919	tcgtatgaagtctgtgtttaaagtgctctatgtcatgagccacaactacaacacatccgagt	1978
Db	1945	tcctatgaaggtttttgttcaaggctctactccatccacacacactatcacatttcggtg	2004
OY	1979	ccagttggccaacttcggtgaagaagagatcatcagatgtgaagtgtgcgaacaagctgtgctcg	2038
Db	2005	ccgggtacctgcctcgtggaagaagatcatcagatgtgacatgacatgacaaacttggtctcaggg	2064
OY	2039	gaaggcctgatcatagtaacagatgagttcccggaaggaaaggtgtgtgtctaaacctaat	2098
Db	2065	gaagccctgtatcatcgttcaagagctctactccatccacacacactatcacatttcggtg	2124
OY	2099	gagtcttcagatattgaagagcgcacacataatgaagcgtgtgtgtgtgtgtgtgtgtgtgt	2158
Db	2125	gattgtttcagatattgtacagacgtctacacatttaattgacagcctgtttgtcctgcccgagag	2184
OY	2159	caattcagatctactgactccctctaccagaacagaaagcccaactgtgtgaacaagtgtga	2218
Db	2185	caattcagatctactgactccctctaccagaacagaaagcccaactgtgtgaacaagtgtga	2244
OY	2219	acttttaactgtagtgcctccaaagaatttagataccagatcacattatgatgttgaa	2278
Db	2245	acattttgagctgatgactctgaanaagcctggcgctacacatatacactacagatttgaa	2304
OY	2279	ctcttcaactgtgcgtacgtgacgtcgtgaactatcatcacacatttgaaggacataattt	2338
Db	2305	ctctttcaactgtgcgtacgtgacgtcgtgaactatcatcacacatttgaaggacataattt	2364
OY	2339	aaaaagaccacacagaacacttgattgttccctgtaggagatattaatgaatcagtttgtg	2388
Db	2365	aaaaagaccacacagaacacttgattgttccctgtaggagatattaatgaatcagtttgtg	2424
OY	2399	gtctcaactgtgacgtcgtccttgttctccagctcaagaacggtgtcagctatitaaaaaa	2458
Db	2425	gttctcactgtgacgtcgtccttgttctccagctcaagaacggtgtcagctatitgaaaaa	2484
OY	2459	tttatataagatagcagccacactgtgaaagatataaaactctgaattccctttttgcaat	2518
Db	2485	tttttttcaagatgagcgtctcactccagagagatcaaaaaactttaaattcccttttgcctatc	2544
OY	2519	gtatcagagcctaataacatttgtctgtagcgtgtgtgacgtacagctgtggagaagaactgca	2578
Db	2545	gtcatggagactcacacaaacgtggccgtagacgctgttg6cactaaacgtggagaaactgtccg	2604
OY	2579	agcaagttcaagaagttcttatgtcggaagtttgaagtttliatgatgaccccttcaagaagcac	2638
Db	2605	agcaagttttagaagttcttatgtcggaagtttgaagtttliatgatgaccccttcaagaagcac	2664
OY	2639	agggcctacaggtctgaacagttagcttgaaacctcctctcatccctccttctgtccttgg	2688
Db	2665	agggcctacaggtctgaacagttagcttgaaacctcctctcatccctccttctgtccttgg	2724
OY	2699	ctcattaaagaatatgacatttactcatgaaagggaaacaagcttcatgtacaatatagta	2758
Db	2725	cttattttaaagttatgacattttactcatgaaagggaaacaagcttcatgtacaatatagta	2784
OY	2759	aactttgaaaaaatgtcgatgatgttgaatatacggccagaaacgtgagatatacagaagc	2818
Db	2785	aactttgaaaaaaaagccatgatgttgcacaaacgtccacgaacagttacgtgttacttaccgagc	2844

QY	2819	caaccctcaatcctgatgacgctcaagctataagaacacacacagatgctcggaagttat	2878
Db	2845	CGAGCTTCATCCGGAGCCGCTCAACCTATATAAGAACCATCCAGGAGTCCGGAGTTAT	2904
QY	2879	gtacgacgaatgaatgibatgttgacacacagagaacattatcacagatgcacagattat	2938
Db	2905	GTACGGCATTTAATGATGATTGACACACAGAGAACTTTATCACAGATGTCACACAGATTA	2964
QY	2939	gagcctcgacacataag	2956
Db	2965	GAGCCTCGAGGCGCATAG	2982
RESULT	3		
LOCUS	RNU78517		
DEFINITION	RNU78517	2399 bp	mRNA
ACCESSION	II (CAMP-GEFII) mRNA, partial cds.		
VERSION	U78517		
KEYWORDS	U78517.1	GI:4115910	
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 2399)		
TITLE	Kawasaki, H., Springett, G. M., Mochizuki, N., Toki, S., Nakaya, M., Matsuda, M., Housman, D. E. and Graybiel, A. M.		
REFERENCE	A family of CAMP-binding proteins that directly activate rap1		
MEDLINE	Science 282 (5397), 2275-2279 (1998)		
REFERENCE	99074384		
AUTHORS	2 (bases 1 to 2399)		
TITLE	Springett, G. M., Kawasaki, H., Graybiel, A. M. and Housman, D. E.		
JOURNAL	Direct Submission		
COMMENT	Submitted (18-NOV-1996) Center for Cancer Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E17-540, Cambridge, MA 02135, USA		
FEATURES	On Jan 8, 1999 this sequence version replaced gi:4106961.		
SOURCE	Location/Qualifiers		
gene	1. .2399		
CDS	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	<1. .2399		
	/gene="CAMP-GEFII"		
	<1. .1311		
	/gene="CAMP-GEFII"		
	/note="similar to Homo sapiens CAMP-GEFII"		
	/codon_start=1		
	/product="CAMP-regulated guanine nucleotide exchange factor II"		
	/protein_id="AAD03423.1"		
	/db_xref="gi:4115911"		
	/translation="RVRLIVLQMAAMYGDLIQEDDVAMAFLEEFVYSVSDARMMVAEKEOLAELEKTVKOISEDAKAPKHKVYLLOOFNGDEPAQKROPIRGSDEVLFPVYCGDITPTTTPVPAASVKEIVASADKISGSELTITVKNNGSEKVIKPNDSVPTTLTINGRIFACPRGDFSLPLPLPQEBGTTGTVTFFELMSKDLATQMTTYDPLPNCVLELELIVFACRNEFNRKFTYANLDLFLRRENIQFVWVTEICLSQSRVQLKQIKITAEACHEKKNLNSFPIVGLISNVAESRLATWEKILPSKKPYAFESLMDPSNRHKAA RUTKAELEPPLIPMLPILKIDMTTHEGNKIFINLVFEKMRMIANTARIVYRSQSPFNDAAGANKNHODVRSYVOLANVINDQPTLSQMSHLEPRRP"		
BASE COUNT	707 a	517 c	522 g
ORIGIN			653 t
Query Match	36.9%;	Score 1112.8;	DB 12; Length 2399;
Best Local Similarity	88.5%;	Pred. No. 0;	
Matches 1207; Conservative	0;	Mismatches 157;	Indels 0; Gaps 0;
QY	1646	cgagtcacgccttgcttctacatgagcgtccatgatatgagacctcctgcaagaagat	1705
Db	1	CGGGTCATCCGCTGGTTCGCAATGGCGAGCAAGTATGATGATCTTCCACAGAAGAC	60

QY 1706 gacgataatgagcccttcctggagagatttatactatcatgatatcatgatatcccgatg 1765
 Db 61 GACGTGGCAGGCGTTTCTTGAGAGAGTTTATGTGTGTATGATGATGATGACGAGATG 120
 QY 1766 attgctgcccccaagagagcaactccagagtttgagaagaattgtcaagcaaatctcga 1825
 Db 121 ATGGTCGCTTTCAGAGACACCTGCTGAGCTGAGAGAGACTGTCAAAACATCTCAGAA 180
 QY 1826 gatgcaagagcaccacaagaagcacaaggttcttttgcaacagttcaatagggagat 1885
 Db 181 GACGCAAAAGCTCCACAGAAAGACACAGAGCTGCTTTTGACGAGTTTAATACAGGTGAC 240
 QY 1886 gagagagcccaagagcgcagcactatccgcgcctctgatgaagttctgtttaagttcat 1945
 Db 241 GAGAGGCGCCAGAAAGCGTCAGCCGATTCGTGTGTGTAAGTTTGTTCAGAGCTTAC 300
 QY 1946 tgcattgagccacactatacaacacattccggtgcgaatgagccacttcgtgaagaagtc 2005
 Db 301 TGCATTGACACACCGATGACACCATCCGAGTGCCTGTGAGCGCGCTGCTGTAAGAGAGTC 360
 QY 2006 atcagttgagttgcgcagcaagcttggtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2065
 Db 361 ATCAGTGCAGTAGACAGACAACTTGCGCTGTGGAGAGGCGCTGATCATTGTCAAGATGAC 420
 QY 2066 tccggaagagaaaggtgtgtctcaaacctaatgtatttcaagatttaagcagctcacc 2125
 Db 421 TCTGAGAGAGAAAGTGTTGCTCAAAACCTATGATGTTCAGATTTAGAGCGCTTACC 480
 QY 2126 attaatgagccttggttctgtctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2185
 Db 481 ATTAAATGACGCGCTGTTGCTGCTGCGCCGAGAGAAATTCGACTACGACTCCCTTGCCA 540
 QY 2186 gaacagagagcgcacactggttgagacagttggaacatttgaactgagctcccaaat 2245
 Db 541 GAACAGAGAGGCGCCGACCACTGCGAGCAGTGGGAGCGTTTGAACGATGAGCTCGAAGAC 600
 QY 2246 ttagcataccagatgacaattatgatttgggaactcttcaactgctgcaatgagctgag 2305
 Db 601 TTGGCGTACCGAGATGACAACTATGACTGGGAACTTTCAACTGTGTGCTCGAGCTGAG 660
 QY 2306 ctactatcatcacatttggagagcacaattttaaaagaccacagcaaaacttggattg 2365
 Db 661 CTAAATCAACACACATTTGGAAAGCATTAATTTAAAGAACACAGCAAACTTGATTTG 720
 QY 2366 ttcttgagaggaatttaagaaattcagtttgggtgctgaactgagatctgcttcttct 2425
 Db 721 TTCTTGAGAGATTTAATGAAATTCAGTTTGGGTTGTCTGATGATCTGCTTTGTGTC 780
 QY 2426 cagctcagcaagcgtgttcagctataaaaaaatltaagaatgagcccaactgtaag 2485
 Db 781 CAGCTCAGCAAGCGTGTTCAGCTTTTGAAAAATGTATCAAGATAGCGGCTCACTGCAAG 840
 QY 2486 gagataaaaaatcgaattccttttttgcataatgctgagactgaagtaacattgctgtg 2545
 Db 841 GAGTACAAAAAATTGAATCTTCTTCTTGGCATGCTCATGGGCTCACTAAGCTGAGCTGAG 900
 QY 2546 agccgcttgagcaactggtgagaacatgccaagcaagtccaagaagtcttcttctgagag 2605
 Db 901 AGCGCGCTGGCATTAATATGGGAAATCTCCGACGCAAGTTTAAAGATTTCTATGGGAG 960
 QY 2606 ttgagaagttaatgagccttcaaggaacacagagcctacagcagctgacagtaagtaag 2665
 Db 961 TTGTAGAGCTTAAATGATCTTCCAGAAATCACAAAGCGTACAGCGCTGACAGCAAGCTAA 1020
 QY 2666 ctggaacctctctcatccctctcatagcttctgctcaattaaagatatgacattactcat 2725
 Db 1021 CTGGAGCGCCCGCTCATCCCTTTCATGCGCTTCTCTTAATTAATAATATGACATTTACTCAT 1080
 QY 2726 gaagggagacaagcgttctgacaactctagtaaaccttgaagaaatgacagatgagta 2785
 Db 1081 GAAGGGAGCAAGCAATTCATTCATGACATCTGATTAACCTTTGAAAAAATGGCGCATGATTGCA 1140
 QY 2786 aatacggccagaaagtgagatactacaagagccaaaccttcaactcgtgatgagctcaa 2845

Db 1141 AATACGCCGAAGCGTGGCGCTACTACAGAGGACCACTTCAATCGGATGCTGCTCAA 1200
 QY 2846 gctataagaacacataagatgtccgaggtatgtatcagcgaatlaatgtatgataaac 2905
 Db 1201 GCTAATAAGAACCATCAGATGTCGCGAGTTATGTACGCGCAATTAATGTGATTGCAAC 1260
 QY 2906 cagaagaattatcacagatgtcacacagattagagccttcgcagcaatagacattcaa 2965
 Db 1261 CAGAGAACTTTATCACAGATGTACACAGATTAGACCTTCGCGCAATGACATCTGCA 1320
 QY 2966 atgcccaagcaagctgttctccagtcacacatttcaaaaa 3009
 Db 1321 GTGCCAGAGTGAATGCTCGCTCCAGTCCACAAATCTTTCAAA 1364
 RESULT 4
 HSU78168
 LOCUS
 DEFINITION Homo sapiens cAMP-regulated guanine nucleotide exchange factor 1
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 TITLES
 AUTHORS
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLES
 JOURNAL
 1 (bases 1 to 3394)
 Kawasaki, H., Springett, G.M., Mochizuki, N., Toki, S., Nakaya, M.,
 Matsuda, M., Housman, D.E. and Graybiel, A.M.
 A family of cAMP-binding proteins that directly activate Rap1
 Science 282 (5397), 2275-2279 (1998)
 2 (bases 1 to 3394)
 Kawasaki, H., Housman, D.E. and Graybiel, A.M.
 Direct Submision
 Submitted (14-NOV-1996) Center for Cancer Research/Department of
 Brain and Cognitive Sciences, Massachusetts Institute of
 Technology, 77 Massachusetts Avenue, E17-540, Cambridge, MA 02135,
 USA
 FEATURES
 source
 Location/Qualifiers
 1..3394
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12"
 /map="12q13"
 1..3394
 /gene="cAMP-GEFI"
 216..2861
 /gene="cAMP-GEFI"
 /note="GEF domain shows highest similarity to RasGEF
 family; contains cyclic AMP binding domain"
 /codon_start=1
 /product="cAMP-regulated guanine nucleotide exchange
 factor 1"
 /protein_id="AA012740.1"
 /db_xref="GI:4079649"
 /translation="MVLRRMRHPPSCSYVLLLEHQHSPCTIGLRMPPLTNSSELSLPS
 ESLSDASTERVLRAGROLQHILATPCNLIRDRKYHLRLROCCSGHELDGILALGL
 GVHSRQYVGCIOVLDEGALCHVKHMAERODRAGYRFPGEPEPVGHHEBELA
 EAVALLSQRGPDALLTVALKRPQGRDEDELIFEEILIKVAHLSNVRLELNG
 LLFPHSKAGVULESQDQKTSWYTIKGSVYNTYTRKGLVTLTHESDDGQALAVND
 APRATILIRETNHFLKVDKQDENRIIKQVEATMRLEHGKVVLYLVERASQAGS
 RPPGRRHRYVMGSPDKLLELLEAMGLDSSAHDKETFLSDFLTHRVFMPSQL
 CAALLHFRHVEPAGSEDERSTYVCKNRQOILRLVSQVALYSGMLTHDVPATSLK
 LSDLVGDRITSLNLRQMPERRRCCHLENGCGNAPOMKARMLPYWLPRQDRELGS
 SCATQVGDVYDLCRPDHSVLTLOLPVTASVEVMAALAEQDQWKGVLYLVNSAG
 DATLODPAKAVATSLINERLRYVNOQYEHLLPHPDQIGPTVGSABEGDLVASOL
 AGQILDHDSLFNSIHQVELIHVYLGQHLRDVTYANLEKRMRENELOLWATVELCL
 CPVGPRAQOLKRFIKLAHLIKDKVNSFFAVMFGLSNPISRLAHTWERLPHKVKR
 LYSALERLLDPSNMHRYRLALAKILSPVLYFPMFLDKWTFLHEGNHTLVENLIME
 KRMMAARAARLHHCRRSHNPVLPSPLSRVSHLHDSQVARISTCSGQSLSTSPAST

BASE COUNT 721 a 996 c 1042 g 635 t
ORIGIN MAYVOLKVIDNORELSRLSRELBP"

Query Match 23.9%; Score 721.6; DB 40; Length 3394;
Best Local Similarity 58.6%; Pred. No. 1.2e-193;
Matches 1359; Conservative 0; Mismatches 934; Indels 27; Gaps 5;

QY 510 aggtcccttcagagaagatccttcagagctggaataatttcaagaatgcattctctcc 569
DB 361 AGGCTCTCACAGAGCGGGTCTCAGGGGCGAGCGAGCTGCATCGCATCTCGGCA 420
QY 570 gaacacctcaatgaagaagataaataacacctaaagaatacagacaatgctg 629
DB 421 CCGGCCAAMACCTCATCCAGACCGGAAGTACCACTTAGGCTCTATCGGAGTGTGCT 480
QY 630 tgggaactgaactggtgagctgagatgacgagacacacatggttcactccgagactc 689
DB 481 CTGGCGGGAGCTGTGGATGGGATCTTGCCCTGGGACTTGGGGTCCATWCCGGAGGC 540
QY 690 aagctgttgcaatgagcaagtcgttgaagaatggttctcaaccaatggaacag 749
DB 541 AAGTTGTGGAACTCTCCAGGTGCTGATGAAGTGCCTCTGCACTGTAACACAG 600
QY 750 agcaacattccaagaatttattatctatcgattctgatagtatgagacagagatg 809
DB 601 ACTGGGCTCTCCAGAGCCAGAT---GCCCAATTCTACCGGTTCCCGGCGCCGAGCCGC 657
QY 810 ccccttgcctactgaagaagaagaagagtgatgataagagctccagacacatgc 869
DB 658 AGCCCGTGGGAATCATGAGATGGAGAGAGATTGGCCAG-----CTGTGG 705
QY 870 tgcgtgttcaacagatgagcccgagcccaatgagatgactctgcgaaccaactg 929
DB 706 CCTGTGCTCCACGCGGGGCTCGACGCCCTGCTCAGTGGCACTTCCAAAGCCCCAG 765
QY 930 gccagagagctgtgataagactaagattatctataagagacttcttaataagact 989
DB 766 GTACAGGCCAGGATGAAGAGCTGGACCTTCTTAGAGAGCTGCTGCACTACAGAGCTG 825
QY 990 tatcccatcttctacacagtgaaacagagtgtaagctgttctcatcttttgaagctc 1049
DB 826 TGGCCACCTCTCCAACTGGGTGAAGCGAATATTAGCGGCTGTTCGCTTTGAACAC 885
QY 1050 aagcccaagagagagctgtgtttaaaccagggggaagaagatcctctgttaca 1109
DB 886 ACAGCAAGGAGGAGCGAGCTGTGTTCTCAGCAGGCGGACAAAGGCACTTGTGTACATTA 945
QY 1110 ttctaaagatcagtgatgtatgattcttaacgcaagggtgtgtctgacacctgcatg 1169
DB 946 TCTGGAAAGGATCTGTCAACGTGTGATCCCATGGCAAGGGGCTGGACCACTTCATG 1005
QY 1170 aaggaatgacttcgcaagtgtagcaactagtgatgaccccaagctgctctatag 1229
DB 1006 AGGAGATGATTTTGGACAGCTGGCTGTGGTAATGATGACACCCCGGAGCCACCATCA 1065
QY 1230 tcttaagagaagataactgccattcttaagagtagaacaagaagatlttaacacgata 1289
DB 1066 TCTGTGGAGTAATCAACTGTCAATTTCTGTGTGACAAAGCAGGACTTCAACCCGTATA 1125
QY 1290 taaggagctgagaggaataacagctcaagcttaagaataatgacaaatgctgtgttc 1349
DB 1126 TCAAGGATGTGGAGCAAAACATCGGCTGGAAGAAACATGGCAAAAGTGTGTGGTGC 1185
QY 1350 tggagaaggtcccaagaggaacagagcttctlatcaaggaactcaacagctcaagca 1409
DB 1186 TGGAGAGAGCTCTCAGGGGGCGGCTTCCCGACCCCAACCCCAAGGCAAGGAACCG-- 1244
QY 1410 agtatctgtatgtcaagaacaactgaanaaattttagagcaatttctagaanaactc 1469
DB 1244 -GTATACAGTATGTCTGGACTCCAGATAGATCCTTAGAGGCTTGTGTGGAGGCGCATGG 1302

QY 1470 gctttgagcaactttaaataagcaacagatctctgtttaaactgaatttataatg 1529
DB 1303 GACTAGATTCCAGTCTCATGACCCAAAAGAAATTCCTTAGGAGCTCTCTCTGACCC 1362
QY 1530 actgtgttttattgacaaataaccaagctttgcccggagactggtgagccactacaagca 1589
DB 1363 ACAGAGCTTATGATGCCAGGCCCAACTCTGCGTGCCTTCTGACCACTTCCATGTGG 1422
QY 1590 agccttcaagagtatacagaagaagaataatgattatgaccttaacaaataagagcgag 1649
DB 1423 AGCTGCGGGTGGCAGCAGAGAGCCAGCACTTACCTGCTGCAACAGAGGAGCA 1482
QY 1650 tcatccgctgtgttcaagtgaggctgcatgatagtgagactctctcaagaagatgag 1709
DB 1483 TCTTGGGCTGTGTACGCCAGTGGGTGGTGGCTGTATGGCTCATGCTCCACACTACCTG 1542
QY 1710 tatctatgacctctcgtgagagatttataatgataatgataatgagtcgagatg 1769
DB 1543 TGGCCACAGCTTCTTCCAAACTCTCAGACCTGTGGGCGAGGAGACCCGACTCAGCA 1602
QY 1770 ctgacctaaagagagactgccaagattgag-----aagattgcaagaactctcag 1823
DB 1603 ACTGCTAGGGAGCAGTGGCCAGAGAGCGGCATGCCACAGGTTGAGATGGCTGTG 1662
QY 1824 aagatgcaaaagcaccacaaagaagcaaggttcttlttcaacagttcaatacgag 1883
DB 1663 GGAATGCATCTCTCAGATGAAGAGCCCGGAACCTTGCTTGTGGCTCCCAACCAAGAGC 1722
QY 1884 atgagaagcccaagaagcgccagactcactcggcgctctgataagattctgtttaaagct 1943
DB 1723 AGCCCTCTCTGAGCAGCTGTGGCATCAAGTTGGGATAAAGTCCCTATGACATCT 1782
QY 1944 attgatatgacacacactcaacaacatctgggtgagcagtgccacttgggtgaagaag 2003
DB 1783 GCCGCGACAGCCACTCACTGATTGTAACCTGCAAGCTCGCTGTACAGCTCTCTGAGAGG 1842
QY 2004 tcatcagtcaggttgcgacaagctgggtctccggggaagggcgtatcatagtaagatga 2063
DB 1843 TGATGGCAGCGTGGGCCCAAGAGATGAGCTTGACCAAGGAGAGTGTGTGAAGCTCA 1902
QY 2064 gtccgagaggaagaagtggtgtctaaacctaaatgatttcaagatlttaagagctca 2123
DB 1903 ATTTCAGAGGTGATCCATTGGGCTGCGACCAAGTGGCCGTTGTGGCCACATCTGTGG 1962
QY 2124 ccaatgaatgagacgctgttcttgcgcgagagcaatctgattcaactgactcccttac 2183
DB 1963 GGCTCAATGAGGTCTCTTGTGTGTAACCCCAAGGAAGTCAATGATGCCACACCC 2022
QY 2184 cagaacaggaagggccaactgttgaacagatgggaactttgaactgatagtctcaag 2243
DB 2023 CTGACCCAGCTGGGGCCCACTGTGGCTTGTGAGGGGCTTGACTGTGATGCCAAGC 2082
QY 2244 attlagcaccaatgacaatttataatgttgaacttcaacttcgltgataagctg 2303
DB 2083 ACCGTGACGGCCAGTGCAGGACGACAGCTGAGGCTCTTCAACAGTATCCACACAGGTGG 2142
QY 2304 agctaatctatacaacatttgaaggaataattt---taaaagacacagacaacttgg 2360
DB 2143 AGCTATCTCACTATGTGTGGGCCCCCGACATCTGCGGGATGTGACCAACCCCACTGG 2202
QY 2361 attgtctcagagagatttaataatcaattgatttgggtcgtacagagatgacctt 2420
DB 2203 AGCGTTATGCGCGCTTTCANATGACTGCACTGAGGTGGGTGCCACCAAGCTGTGTCT 2262
QY 2421 gtctcagctcagcaagcglttcagctatataaataattatlaagaatagcaagccact 2480
DB 2263 GCCCGGTGCCGGGCCCGGCGCCAGTGTCTCAAAAATTCATTAAAGTGGCGCCACCC 2322
QY 2481 gtaagagataaanaatctgatactcttcttgcacatgctatgagagtaagtaactg 2540
DB 2323 TCAAGGACAGAGATGTCAATTCCTTGTGGCGCTGATGTTGTGGCTGACGAACCTGCG 2382
QY 2541 ctgtgagcgcttggcgtataacgttggaggaataatcgtccaagaagttcaagaagtctatg 2600

Db 2383 CCATGACGGCGCTAGCCACACACTGGAGCGGCTGCTCAGCAAAAGTCGGAAGCTGTAAGT 2442
QY 2601 cggagattgaaagtttaatgacaccttaagaacacaagggcctacagctgacagtag 2660
Db 2443 CGGCGCTCGAGAGGCTGCTGCTCATGATCCATGACACCGGGGTATACCGACTGCGCCCG 2502
QY 2661 ctaagctggaacctctcctcatccctcatgacctgctgctcatatgaatagatgacatt 2720
Db 2503 CCAAGCTCTCCCTCTGCTGCTCATGATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 2562
QY 2721 ctcatgagggggaacaacagcttcatatgacaaatcgaatgaatgaatgaatgaatga 2780
Db 2563 TTCAATGAGGGAACACACACACTAGTGAGAAATCATCACTTGAGAAATGAGAAATGA 2622
QY 2781 ttgcaaatagcgccagcaagctgagatactacagagacca 2820
Db 2623 TGGCCAGAGCCGCGGATGCTGCACCACTGCCGAAGCCA 2662

RESULT 5
AF103905 2646 bp mRNA PRI 06-DEC-1998
LOCUS AF103905 Homo sapiens Rapi guanine-nucleotide exchange factor mRNA, complete cds.
DEFINITION AF103905 AF103905.1 GI:3978530
ACCESSION AF103905
VERSION AF103905.1
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2646)
REFERENCE de Rooij,J., Zwartkruis,F.J.T., Verheijen,M.H., Cool,R.H.,
AUTHORS Nijman,S.M., Wittinghofer,A. and Bos,J.L.
TITLE Epac is a Rapi guanine-nucleotide-exchange factor directly
activated by cyclic AMP
JOURNAL Nature 396 (6710), 474-477 (1998)
MEDLINE 99068645
AUTHORS de Rooij,J., Zwartkruis,F.J.T. and Bos,J.L.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) Laboratory for Physiological Chemistry,
Utrecht University, Universiteitsweg 100, Utrecht 3584 CG, The
Netherlands

FEATURES
source Location/Qualifiers
1..2646
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
/tissue_type="muscle"
1..2646
/note="contains cAMP binding domain; Epac"
/codon_start=1
/product="Rapi guanine-nucleotide exchange factor"
/protein_id="AAC83381.1"
/db_xref="GI:3978531"
/translation="MYIRRMHRRRSCTYLLEHQRSCITQGIKMTPLINSESLDPS
ESLEQASTERVLAGRQLHRLHATACENILTRDKYHLRYRQCCSRRELVDGLIAGL
GVHSRSOVVIGICVLLDEGALCHVKHDMAQODDAFYRPPGPEPVTHHEEELAA
EAVALLSQRPDALLVYALRKPGQRDELDLIFELHAIKVAHLSVSKELAAV
LLEPHSKAGTIVLESGDGKSTWYITMKGSVNVYTHGKGLVTHLHGDDGOLATIND
APRAATITLEDNCHLRVDKOPNRIKRYEAKTRBLEHGKVVYLVLEASGAGPS
RPPPGRNRTVMSGPEKLTLELLLEMGDSSADHPTEFLISDFLTHRVKPSQOL
CAALIHFIHVEPAGSGEQERSTVYCNKRQDILRVSOVALYOSMLHTPVPVTSFQK
LSDLVGDRILSLMLEQWERRRCHLEGCNAPOMKARMLPYMLNDEPELGS
SCAIVGDKVPIYICRPDSVLTFLQVTSVEVMAALAOEGWTKGVIVKVNAG
DAIGLOPDARGVATSLGINRLFPVNPQEAHELIIPDOLGPTVGSABEDLVASADL
AGQITBDMSLFNSHQVELIHYLPQHLRDTVTLANLEFMRKHELDIWAETELCL
CPVGPRAQLLRKFIKLAHLKQKMLNSFAVFWFGLSRLAHTWERLPHKVKR
LYSALERLDPNNHRYRLAKLSPPVLPFMPLLKDMTEFHEDGNLVERLLNFE
KRMMAAARMLHHRSHNPVLPSPRSRVSHLEDSSQVARISTCEQSLSTRSPAST

BASE COUNT 543 a 825 c 781 g 497 t
ORIGIN
MAVVOALKVIDNORELSRLRELEP*
Query Match 23.9%; Score 720; DB 40; Length 2646;
Best Local Similarity 58.5%; Pred. No. 3.4e-193;
Matches 1358; Conservative 0; Mismatches 935; Indels 27; Gaps 5;
QY 510 agtgccttcagaagaatcctcagaagctggaanaatttaagaaatgcattcctc 569
Db 146 AGGCTTCACAGAGCGGGCTGCTCAGGGCTGGAGCAGCTGCATCGCATCTCTGGCCA 205
QY 570 gagcaccatcagataagaagatagaanaatcaccaataagaacatcagaatgtgtg 629
Db 206 CTTGCCAATACCTCACCCGAGCGGAAATCACCTTGAAGCTCTTATGCGAGTGTGCT 265
QY 630 tgggaactggaactgtggaactgatatgcagagacacatgtgttcactccggatc 689
Db 266 CTGGCGGAGAGCTGTGATGGGATCTTGCCCTGGGACTTGCGGCTTCATTCGCGAGCC 325
QY 690 aagctgttgcatgtggaagtcgttagaagatgtgttcacacacgttgaccag 749
Db 326 AAGTTGTGGGAATCTGCAGAGTCTGCTGATGAAGTGCCTCTGCATGTGAAACAG 385
QY 750 agcacaattcagaactttattatctatcattctcgtatgtagagcagagatg 809
Db 386 ACTGGGCTTCCAGAGCCAGAT--GCCCAATTCAACGGTTCCCGGGGCCCGAGCCCG 442
QY 810 ccccttgcctactggaagaagaagaagatgtatgtagagatccacagacacatgc 869
Db 443 AGCCCGGAGAACTCATAGATGAGAGAGAGATGTGGCCAG-----CTGTGG 490
QY 870 tctgtgttcaagatgtggcccgacgcacaatgagatgaccttcgcaaacactg 929
Db 491 CCTGTCTCTCCAGCGGGGCTCGAGCGCTGCTCATCTGTGCACTTGCAGAACCCCGAG 550
QY 930 gccaagagactgtgtagtgcctagatatcattagagaagctctcatttaagcct 989
Db 551 GTACGGCACGGATGAAAGAGTGTGACCTCATTTAGAGAGCTGTGCACATCAAGGCTG 610
QY 990 tatccatcttctacacaagtgaacagagatgaagctgtctcatcttttgatctc 1049
Db 611 TGGCCCACTCTCAACTGCGTGAACGAGAAATTACGGCTGTTCGCTTTGAACCC 670
QY 1050 agcacaagaagagactgtgtttaaaccagggggaagaagctcctcgttatcat 1109
Db 671 ACAGCAAGGAGGACCGGTGTGTCAGCCAGGGGACAAAGGCACTTGTGTACATTA 730
QY 1110 ttctaaagatcagatgaatgtatgatttaagcagaaggtgtgtcagaccgtgcatg 1169
Db 731 TCTGGAAAGGATTTGTCAACGTGTGACCCATGTGGCAAGGGCTGTGACACCTGCATG 790
QY 1170 aaggaatgacttcgcaagtttagcactagttgaatgatccccaagagctgctctatcg 1229
Db 791 AGGAGATGATTTTGGACAGCTGGCTCTGTGATGATGTCACCCGGGAGGACACCATCA 850
QY 1230 tcttcaagaagaataactcattcttaagaagtagaagaagatcttaaccagatcc 1289
Db 851 TCTCTGGAGAGCAACACTGTCATTTCTCTGCTGTGACCAAGCAGGACTTCAACCTATCA 910
QY 1290 taaggagacttgaggggaatacagtcagacttaagaanaatagacaagaatgtcgtg 1349
Db 911 TCAAGGATGGGAGGCAAAAGCCATCGGCTGGAAMAATGGAATGAGTGTGTGTGC 970
QY 1350 tgaagaagttcccaagaaggaacagagctcttaactaagaagaacccaagctcaga 1409
Db 971 TGAAGAGAGCTCTCAAGGCGCGGCTTCGCCAGCCCAACCCAGCAGGAACCG-- 1029
QY 1410 agtatactgtatgtcagaagaacacttgaanaaaatttaagatcttctagaacaat 1469
Db 1029 -GTATACATGATGCTGTGACCCCGAGAGAAATCTAGACCTTCTGTGTGAGGCGCATGG 1087

QY	1470	gctttagagcaacttaataagaacaacaaattctgtttaatagcttatatgatgc	1529
Db	1088	GACCAAGTTCACAGGCTCATATACCAAGACAGACATTCTCCAGCCTCTCTCAACC	1147
QY	1530	actgtattatcccaataaccagcttgccgcgactggtgcacactacacgaac	1589
Db	1148	ACAGGGTCTTATATGCCACAGCGCCCACTCTGCGCTGCTTCTGCAACACTTCCATGTGG	1207
QY	1590	agccttcaacaagtacagaaacagagagaanaa tggatlatgacctcaacaaataagagtcgag	1649
Db	1208	AGCCGTGGGGGAGGAGAGAGAGGCGAGCACTACGTACGTCCCAACAGAGGACACAGAA	1267
QY	1650	tcatcgccctggtctcactagttggctgcaatgtatggagaacctctgcagaagatgatgc	1709
Db	1268	TCTTGGCGGCTGTGACGCGAGTGGGTGGCTGTATGGCTCATATGCTCCACACTGACCTGTG	1327
QY	1710	tatctatgacctcttcgtagagatgtatgtatctatcatagatgtatgcccagatgatg	1769
Db	1328	TGGCCACCAAGCTTCTCCAGAACTCTCAGACCTGTGGGAGGAGACCCGACTAGCA	1387
QY	1770	ctgcacctaaagagcaactgcacagatctgag-----aagatgtccaagcaatctcag	1823
Db	1388	ACCTGCTTAGGGAGACAGTGGGCACAGAGGGCGGATGCCACAGATTGGAGAAATGGCTGTGG	1447
QY	1824	aagatgcaaaaggacacacaagaagacagaagttcctttgtcaacagttcaatacagtcg	1883
Db	1448	GGAAATGCATCTCTCAATATGAGGCCCGGAACCTTGCTTTGGCTCCCAACACAGAGAG	1507
QY	1884	atgagagagcccagaagcgcaagcactatccgcgctctgataagttcgtttaagttct	1943
Db	1508	AGCCCCCTTCGGGACAGAGCTGTGCTCCCAAGTGGGGTAAAGTCCCTATGACATCT	1567
QY	1944	atgcattgacaacactatacaaacattcggttgcaatgagcgaactcgtgtgaagaag	2003
Db	1568	GCCGGCCAGACCACTCAAGTGTGACCTGCGACGCTGCTGACAGGCTCCGTAGAGAG	1627
QY	2004	tcatcatgtcagttgtgcgacaagctggtccggtcggtgagggcctgataagtaagata	2063
Db	1628	TGATGGCAGCCTTGGCCAGAGGATGTGACCAAGAGGGGAGTGTGTAAGTCA	1687
QY	2064	gttcgagagggaaaaaggtggtgtcctaactatgatgttcaagratattaagaagctca	2123
Db	1688	ATTTCGACGGGATGACCATGTGGCTGACAGCAAGTGGCCGTGTGTGGCCACTTCTCTGG	1747
QY	2124	ccatlaattgagcgcgtgtgtctgtccgcgagagaacattcgattcaactgacccctta	2183
Db	1748	GGCTCAATGACGCTCTTCTTGTTCACACCAAGAGAGGCGCATGTAGCTATCCACAC	1807
QY	2184	cagaacaggaagccccaactgttggaaacagltggaaacttttgaactgatgagctccaag	2243
Db	1808	CTGACACACTGGGGCCCACTGTGGGCTGTGCTGAGGGGCGTGAACGTGTGAGTGC	1867
QY	2244	attagcataccagaatgaacattatgattgggaactctcaactcgtgcgtgagatgtg	2303
Db	1868	ACCTGGCAGGCGACACTGACGAGCACACAGTGGAGCCCTTTCACACAGTATCCACAGTGG	1927
QY	2304	agcctaactatcacacaatttggaaagcataattt---taaaaagaccacaagaactgtg	2360
Db	1928	AGCTGATTCATTATGTGCTG86CCCCCAGCATCTGCGGGATGTACACACCGCCACACTGG	1987
QY	2361	attgttctcctagagagaatttaataatcaagtatttggcgtgcactgagatctgcctt	2420
Db	1988	AGCGTTTATATGCGCGCTCATATGATGAGTGAAGTACTGAGTGGGCGCACGAGCTGTCTCT	2047
QY	2421	gttctcagctcagcaagcgtgttcagctatataaaaaattatlaagatagcagccact	2480
Db	2048	GCCCGCTCCCGGCGCCCGGCGCCAGCTGCTCAGGAATCATTAAGCTGCGCGCCACAC	2107
QY	2481	gtaaagagatatataaactgaattcccttttttgcacatgcgtcatgaggaactaagttaactg	2540
Db	2108	TCAAGGACCAAAATCTCAATTCTCTTCTTTGGCGGTATGTTTGGCTCCAGCAACTCGG	2167
QY	2541	ctgtgagccgcttgcacatacgttggagaagaactgcgaagcaagcttcaagaattctatg	2600

Db	2168	CCATCAGCCCGCTTAGCCCACTGGAGCGGCTGCCCAACAAGTCCGGAAGCTGTACT	2227
QY	2601	cgaagttgaagttaataagaccttaagaagacaagggcctcaagcgtgacagtg	2660
Db	2228	CCGCGCTCGAGAGGCTCTGTGATCCCTCATGGAACACCGGGTATNACGACTGCGCTTCG	2287
QY	2661	ctaagctggaactcctctcatcccttatgctctgtctatcaataagatatgaattta	2720
Db	2288	CCAAGTCTCCCGCTCGTGTGATCCCTCATGCGCTTCTTCCAAAGACATGACCTTCA	2347
QY	2721	ctcattgaggggaacaagaagttcattgacaatctagtaactttgaaaaaatgcatga	2780
Db	2348	TTCTATGAGGGAACACACACTAGTGGAGATCTCATCAACTTTGAGAAATGAGAAATGA	2407
QY	2781	ttgcaaatagcgccagaagatgagatcactacagggagcca	2820
Db	2408	TGGCCAGAGCGCGGGAGTGTCTGACCACTGCGGAAGCCA	2447

RESULT	6
NUM7167	
LOCUS	NUM7167 3373 bp mRNA ROD 05-FEB-1999
DEFINITION	Rattus norvegicus cAMP-regulated guanine nucleotide exchange factor
ACCESSION	U78167
VERSION	U78167.1 GI:4079656
KEYWORDS	Source
ORFANISM	Norway rat.
REFERENCE	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 3373)
AUTHORS	Kawasaki, H., Springer, G. M., Mochizuki, N., Toki, S., Nakaya, M., Matsuda, M., Housman, D. E. and Graybiel, A. M.
TITLE	A family of cAMP-binding proteins that directly activate Rap1
JOURNAL	Science 282 (5397), 2275-2279 (1998)
MEDLINE	99074384
REFERENCE	2 (bases 1 to 3373)
AUTHORS	Kawasaki, H., Housman, D. E. and Graybiel, A. M.
TITLE	Direct Submission
JOURNAL	Submitted (14-NOV-1996) Center for Cancer Research/Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E17-540, Cambridge, MA 02135, USA

FEATURES	
source	Location/Qualifiers
gene	1..3373 /organism="Rattus norvegicus" /db_xref="taxon:10116"
CDS	1..3373 /gene="cAMP-GEF1" /note="cAMP-GEF1" /note="GEF domain shows highest similarity to RasGEF family: contains cAMP binding domain" /codon_start=1 /product="cAMP-regulated guanine nucleotide exchange factor 1" /protein_id="AAD12739.1" /translation="WVKRMHPRCSQVLFENHRPSCIOGLRWPFLTNSEGLDRVSLDEATPEHLVHAKKGLYHLNLAATPELIDRLRYHRLHROCGSGEINDJTAGLGVHSRQANGTGVQLDLDBALCHVYHMDTPDRDROKPIRFPGPBPQAGNHDEEELVEMALISQNGPDLALVLRKSPGQRTDEIDLFEELVHTKAVAHLSNVKRELAVLLEPHSKAGTVLFEQGDKGSWTIMGVSVAVTYRGKVLTLHEDDVLSGOLAVDAPRAATILIRENNCHFLRYDKDFNRIKIDVETAMRLERKGVVLYLERTSGAGPSRPPTGPNRNYTVMSGTEPELLELLEAPRDSASHDPEFLSDFLTLHVFMPCTQLPALHLHFVEPSEPAQSGEORERTYINRKOOLRLVSRWALYSMLSDPYATSFLOKTSIVSRDRLSNLRAQDPERRRHNRLENCQVSPOTKRNKPYRPHNEPTPSAGATIRGDKVVPIDICRPSVLTLLHLPYASVREYMALEHEDWTGQVLYKYNSSADVVGLQPDARGVATSGLNERLFVVDVEVETLPHRPOGLPTLGSSMDILVSAKIDAGOLETHEDHFNRIHOVELHYVLVGPQHLRVYTLANDERPMRPNELQVWATELCICPVPGRADLLKFKIKLAHTEKQINISFPAVMFGLSNSAISRLATWETLPK

QY	2155	aggcaattcgtatcactgctctccctcaacgagaaagcagcgaactglttgaaacgt	2214
Db	1012	AGACCTGCGGACACTTTGAACCCATTGTGAGAAATTTGAGAAATTCACAGCAAGCAAGGTGCAT	1071
QY	2215	gggaactcttgaaactgtagctgcccacaagaattagcataccagatgaacatttaagctg	2274
Db	1072	GAGGATTTTGGGA---ATGAACAACCTTGGGATCTTGCTGTGGAATTAATGAATTTTGATTG	1128
QY	2275	ggaactcttaactcgtcgctgacgctgagctaaactctatcaacatttggaaggcataa	2334
Db	1129	GAGCTCTTTCATTTCAATTCACAGCAAGAGCGTGATCTACTTCACGTTCCAGCCAGCAAGG	1188
QY	2335	ttttaaaagaccacgaacaaacttgatctgtctccgagagagatttaagtaattcatt	2394
Db	1189	AAGTGGGGAAACACACTGCAAAATTCACGCCCTTCTGCTCCAGAAATGCAATGAAGTCCACT	1248
QY	2395	ttggctgcctactgtagactcgcctctgtctctcaagctcaagcgtgctcaagctataa	2454
Db	1249	TTGGGTGGCCACGGAGATTTGCTGTGACGCCAGTGGGCAAGCGAGTGGACGTGTGTA	1308
QY	2455	aaaaatttaataagatagcagcccaactgtaagaagataaaatctgaattcccttttgc	2514
Db	1309	AAAATTCATCAAAATTCGCGCTACTGCAAGCCGACGAAACCTGAAATTCCTTCTTTCG	1368
QY	2515	catgctcatggagactaagtaacatgctgttgagccgcttgcaactaaqitggagaact	2574
Db	1369	CATTGCTATGGGCTCAACACTCTCTTCTGTCACTGCTGTCAGACTGTGCGACACTGGGAAATAT	1428
QY	2575	gcccaagcaagltcaagaagltcctatgcyggaglttgaaaglttaatgacccctcaagaa	2634
Db	1429	CCCTGGGAATTTAAGAAATCTTTCTGTGAACCTTGAAATTTAACAGATCTTCCCTAA	1488
QY	2635	cccaagggcctcaagcgtgacagtaagctgaagctgcgaacctctctcatccctcttaagcc	2694
Db	1489	TCACAAAAGCCCTACGAGACTCATTTCAAAAAGATGAAGACCCACAAAATTCCTTTATATCC	1548
QY	2695	tttgctcaatlaaagatagctatcttaactactgagggaaacaagacttcatatgaacact	2754
Db	1549	CTTATTCCTTAAGAGTGTATACATTATTTCATGAAGSAAATTAANAACCTTTTGGATAACT	1608
QY	2755	agtaaccttggaaaaaatgcgcatgatlqcaaatcagcgcgaagacagtgaatactacag	2814
Db	1609	TGTCAAATTTTGGAAAAGCTGCTATGTATGATCGACACACTGTCCAAACCTCGAGACACTGAG	1668
QY	2815	gagcaacccctcaactcgtatgcagctgaagctaaagaaacacagagctaggtgtccgag	2874
Db	1669	GACTAACCACTT-----TGGTAACTGTCTCCAAAAGAGCATCAAGACTTAAATCTC	1719
QY	2875	ttatgtacggcaatlaaatgtatgtatgacaacagagaacattatcaagatgtcacag	2934
Db	1720	CTATGTTAAATACCTGATGTATGTATGACAGCCAGCGGCTGTGTTGAGCTCTCACACAG	1779
QY	2935	attagagcctcg 2946	
Db	1780	GATGAGCCTTCG 1791	

RESULT	8
AC016789/c	
LOCUS	
AC016789	59255 bp DNA HTG 05-DEC-1999
DEFINITION	Homo sapiens clone RP11-22F10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC016789
VERSION	AC016789.1 GI:6538791
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE	Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 59255)
TITLE	Birren,B., Linton,L., Nusbaum,C. and Landier,E.
JOURNAL	Homo sapiens, clone RP11-22F10
	Unpublished

REFERENCE
AUTHORS

REFERENCE
AUTHORS
2 (bases 1 to 59255)
Barren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barua, N., Beckerly, R., Boguslavsky, L., Bourkhalter, B.,
Bowie, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cook, P., DeRellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Feirreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karstas, A., Klein, J.,
Lelock, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefaye, S., Tirrell, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (05-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker

Submitted (5-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: I4108

Center clone name: 22_F_10

NOTE: This record contains 76 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1	754:	contig of 754 bp in length
		gap of unknown length
755	1539:	contig of 785 bp in length
		gap of unknown length
1540	2284:	contig of 745 bp in length
		gap of unknown length
2285	3047:	contig of 763 bp in length
		gap of unknown length
3048	3841:	contig of 794 bp in length
		gap of unknown length
3842	4639:	contig of 798 bp in length
		gap of unknown length
4640	5394:	contig of 755 bp in length
		gap of unknown length
5395	6157:	contig of 763 bp in length
		gap of unknown length
6158	6908:	contig of 751 bp in length
		gap of unknown length
6909	7692:	contig of 784 bp in length
		gap of unknown length
7693	8484:	contig of 792 bp in length
		gap of unknown length
8485	9262:	contig of 778 bp in length
		gap of unknown length
9263	10032:	contig of 770 bp in length
		gap of unknown length
10033	10816:	contig of 784 bp in length
		gap of unknown length
10817	11601:	contig of 785 bp in length
		gap of unknown length
11602	12376:	contig of 775 bp in length
		gap of unknown length

*	12377	13153: contig of 777 bp in length
*		gap of unknown length
*	13154	13939: contig of 786 bp in length
*		gap of unknown length
*	13940	14734: contig of 795 bp in length
*		gap of unknown length
*	14735	15510: contig of 776 bp in length
*		gap of unknown length
*	15511	16284: contig of 774 bp in length
*		gap of unknown length
*	16285	17070: contig of 786 bp in length
*		gap of unknown length
*	17071	17843: contig of 773 bp in length
*		gap of unknown length
*	17844	18622: contig of 779 bp in length
*		gap of unknown length
*	18623	19407: contig of 785 bp in length
*		gap of unknown length
*	19408	20209: contig of 802 bp in length
*		gap of unknown length
*	20210	20980: contig of 771 bp in length
*		gap of unknown length
*	20981	21754: contig of 774 bp in length
*		gap of unknown length
*	21755	22514: contig of 760 bp in length
*		gap of unknown length
*	22515	23302: contig of 788 bp in length
*		gap of unknown length
*	23303	24106: contig of 804 bp in length
*		gap of unknown length
*	24107	24874: contig of 768 bp in length
*		gap of unknown length
*	24875	25638: contig of 764 bp in length
*		gap of unknown length
*	25639	26379: contig of 741 bp in length
*		gap of unknown length
*	26380	27171: contig of 792 bp in length
*		gap of unknown length
*	27172	27954: contig of 783 bp in length
*		gap of unknown length
*	27955	28728: contig of 774 bp in length
*		gap of unknown length
*	28729	29507: contig of 779 bp in length
*		gap of unknown length
*	29508	30299: contig of 792 bp in length
*		gap of unknown length
*	30300	31100: contig of 801 bp in length
*		gap of unknown length
*	31101	31880: contig of 780 bp in length
*		gap of unknown length
*	31881	32636: contig of 756 bp in length
*		gap of unknown length
*	32637	33420: contig of 784 bp in length
*		gap of unknown length
*	33421	34184: contig of 764 bp in length
*		gap of unknown length
*	34185	34973: contig of 789 bp in length
*		gap of unknown length
*	34974	35763: contig of 790 bp in length
*		gap of unknown length
*	35764	36547: contig of 784 bp in length
*		gap of unknown length
*	36548	37311: contig of 764 bp in length
*		gap of unknown length
*	37312	38101: contig of 790 bp in length
*		gap of unknown length
*	38102	38894: contig of 793 bp in length
*		gap of unknown length
*	38895	39670: contig of 776 bp in length
*		gap of unknown length
*	39671	40443: contig of 773 bp in length
*		gap of unknown length
*	40444	41210: contig of 767 bp in length

*	41211	41988: contig of 778 bp in length
*		gap of unknown length
*	41989	42772: contig of 784 bp in length
*		gap of unknown length
*	42773	43548: contig of 776 bp in length
*		gap of unknown length
*	43549	44334: contig of 786 bp in length
*		gap of unknown length
*	44335	45138: contig of 804 bp in length
*		gap of unknown length
*	45139	45923: contig of 785 bp in length
*		gap of unknown length
*	45924	46706: contig of 783 bp in length
*		gap of unknown length
*	46707	47484: contig of 778 bp in length
*		gap of unknown length
*	47485	48254: contig of 770 bp in length
*		gap of unknown length
*	48255	49016: contig of 762 bp in length
*		gap of unknown length
*	49017	49810: contig of 794 bp in length
*		gap of unknown length
*	49811	50603: contig of 793 bp in length
*		gap of unknown length
*	50604	51376: contig of 773 bp in length
*		gap of unknown length
*	51377	52156: contig of 780 bp in length
*		gap of unknown length
*	52157	52936: contig of 780 bp in length
*		gap of unknown length
*	52937	53728: contig of 792 bp in length
*		gap of unknown length
*	53729	54498: contig of 770 bp in length
*		gap of unknown length
*	54499	55296: contig of 798 bp in length
*		gap of unknown length
*	55297	56090: contig of 794 bp in length
*		gap of unknown length
*	56091	56878: contig of 788 bp in length
*		gap of unknown length

Query Match
Best Local Similarity 5.8%; Score 176.2; DB 44; Length 59255;
Matches 178; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	2829	atcctgagtcagctcaagctaataagaaccatcagatgtccgagttatgtacgcaat	2888
Db	46315	ATCCTGAGTCAGCCTCAGAGCTATTAAGAACATTCGATGTCGAGTTATGTACGGCAAT	46256
QY	2889	taaatgtgattgacaacccaggaaccttatccacagatgtcacacagattagagcctcgtc	2948
Db	46255	TAAATGTGATTTGACACACAGAGACTTTATTCACAGATGTACACAGATTTAGAGCCTCGTC	46196
QY	2949	gaccatgagaccttcaaatgcccacaagaagttgtctcccgatccacaattttcaaaa	3008
Db	46195	GACCATGACATTTTCAAATGGCCAAAGCAACAGTTTGTCCTCCATCCACATTTTCAAA	46136
QY	3009	a	3009
Db	46135	A	46135

RESULT 9
AC014497
LOCUS
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014497
VERSION AC014497.1 GI:6436838
KEYWORDS HTG; HTGS PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

SOURCE	ORGANISM	fruit fly, Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	AUTHORS	1 (bases 1 to 189668) Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazek,R.G., Dolnikoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequelra,A., Sethi,H., Snir,E., Sylrskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE	JOURNAL	Sequencing of Drosophila melanogaster
REFERENCE	AUTHORS	Unpublished 2 (bases 1 to 189668) Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazek,R.G., Dolnikoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequelra,A., Sethi,H., Snir,E., Sylrskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE	JOURNAL	Direct Submission
COMMENT		Submitted (21-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA On Aug 2, 1999 this sequence version replaced gi:4887256. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdg@fruitfly.berkeley.edu . All contigs in this submission meet the following cutoffs: length >= 200 bases. * NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
		1 586: contig of 586 bp in length * 587 566: gap of unknown length * 667 1822: contig of 1156 bp in length * 1823 1902: gap of unknown length * 1903 2645: contig of 743 bp in length * 2646 2725: gap of unknown length * 2726 3326: contig of 601 bp in length * 3327 3406: gap of unknown length * 3407 4154: contig of 748 bp in length * 4155 4234: gap of unknown length * 4235 5204: contig of 970 bp in length * 5205 5284: gap of unknown length * 5285 6888: contig of 1604 bp in length * 6889 6969: gap of unknown length * 6970 8402: contig of 1434 bp in length * 8403 8482: gap of unknown length * 8483 10183: contig of 1701 bp in length * 10184 10263: gap of unknown length * 10264 11736: contig of 1473 bp in length * 11737 11816: gap of unknown length * 11817 13661: contig of 1845 bp in length * 13662 13741: gap of unknown length * 13742 15560: contig of 1819 bp in length * 15561 15640: gap of unknown length * 15641 17739: contig of 2099 bp in length * 17740 17819: gap of unknown length * 17820 19768: contig of 1949 bp in length * 19769 19848: gap of unknown length * 19849 23150: contig of 3302 bp in length
		23151 2330: gap of unknown length * 23231 26360: contig of 3130 bp in length * 26361 26440: gap of unknown length * 26441 32462: contig of 6022 bp in length * 32463 32542: gap of unknown length * 32543 38690: contig of 6147 bp in length * 38690 38779: gap of unknown length * 38779 44143: contig of 5374 bp in length * 44144 44224: gap of unknown length * 44224 53717: contig of 9493 bp in length * 53717 53796: gap of unknown length * 53796 60200: contig of 6404 bp in length * 60200 60280: gap of unknown length * 60281 68813: contig of 8533 bp in length * 68813 68893: gap of unknown length * 68893 82418: contig of 13525 bp in length * 82418 82419: gap of unknown length * 82419 82499: gap of unknown length * 82499 110321: contig of 27823 bp in length * 110321 110401: gap of unknown length * 110402 133615: contig of 23214 bp in length * 133616 133695: gap of unknown length * 133696 168863: contig of 35168 bp in length * 168863 168943: gap of unknown length * 168943 169169: contig of 226 bp in length * 169169 169249: gap of unknown length * 169249 169250: gap of unknown length * 169250 169984: contig of 734 bp in length * 169984 170063: gap of unknown length * 170063 170787: contig of 724 bp in length * 170787 170867: gap of unknown length * 170867 171611: contig of 694 bp in length * 171611 171642: gap of unknown length * 171642 172348: contig of 707 bp in length * 172348 172429: gap of unknown length * 172429 173135: contig of 707 bp in length * 173136 173215: gap of unknown length * 173215 173826: contig of 611 bp in length * 173826 173906: gap of unknown length * 173906 174615: contig of 709 bp in length * 174615 174695: gap of unknown length * 174695 175429: contig of 734 bp in length * 175429 175509: gap of unknown length * 175509 176117: contig of 608 bp in length * 176117 176197: gap of unknown length * 176197 176118: gap of unknown length * 176118 176892: contig of 695 bp in length * 176892 176971: gap of unknown length * 176971 176972: gap of unknown length * 176972 177591: contig of 619 bp in length * 177591 177671: gap of unknown length * 177671 178328: contig of 657 bp in length * 178328 178408: gap of unknown length * 178408 179175: contig of 767 bp in length * 179175 179255: gap of unknown length * 179255 179588: contig of 603 bp in length * 179588 179938: gap of unknown length * 179938 180604: contig of 666 bp in length * 180604 180684: gap of unknown length * 180684 181325: contig of 641 bp in length * 181325 181405: gap of unknown length * 181405 182142: contig of 737 bp in length * 182142 182222: gap of unknown length * 182222 182948: contig of 726 bp in length * 182948 183028: gap of unknown length * 183028 183685: contig of 657 bp in length * 183685 183765: gap of unknown length * 183765 184460: contig of 695 bp in length * 184460 184540: gap of unknown length * 184540 185146: contig of 606 bp in length * 185146 185226: gap of unknown length * 185226 185843: contig of 617 bp in length * 185843 185923: gap of unknown length * 185923 186753: contig of 752 bp in length * 186753 186755: gap of unknown length * 186755 187433: contig of 678 bp in length * 187433 187513: gap of unknown length


```

FEATURES
  source
    1..189668
      /organism="Drosophila melanogaster"
      /strain="y: cn bw sp"
      /db_xref="taxon:7227"
      /chromosome="2"
      /map="42E-43A"
      /clone="BACR10F15 (D621) RPCI-98 10.F.15"
      /clone.lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACs.6"

BASE COUNT   51039 a 41017 c 40764 g 52607 t 4241 others

ORIGIN
Query Match      4.7%; Score 141.8; DB 33; Length 189668;
Best Local Similarity 65.8%; Pred. No. 9.2e-29;
Matches 206; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY  2479 ctgtaaaggagataaaatctgaattccttttggccatcgatcgagactaagtaacat 2538
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  28022 CTCGAAGAGGATACAGACACTTAACGCCCTTCTTCGTGGTAATGCGTCTCCACACT 27963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  2539 tgcgtgagcgcgttgacactaagctgggagaagaactgccaagaattcaagaattcta 2598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  27962 GGGGGTGTCCAGGCTGCACAGACACTGGGAAAAGATTCCGTGAGATTGAGAGATCTT 27903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  2599 tgcggagtttgaagttaattgaccttcaaggaacccaagagcctacaagctacagat 2658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  27902 CCAAGGAGTTGAGGGCCCTGATGACCCAGTCGCAACCAAGGCGGTACCGAGTGTCT 27843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  2659 agctaaagctggaaacctccctcatccctcatccctgaccttgcctatgaagatatacatt 2718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  27842 GGGCAAGCTGCACGCCCGCTTGATTCCTTCATGCCGCTCCTGCTCAAGAAATGACCTT 27783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  2719 tactcatggggggaagaagacttcatgacatactagtaaaacttgaaaaaatgcgcat 2778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  27782 CGGCCATGAGGGGCAACAGACAGCCTGAGCGGCTGTGAACTTGAGAAATGACACT 27723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  2779 gattgcaataagc 2791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  27722 GATGGCCAGACG 27710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AC008340/c 125681 bp DNA HTG 06-AUG-1999
LOCUS      Drosophila melanogaster chromosome 2 clone BACR07420 (D918) RPCI-98
DEFINITION 07.J.20 map 42D-42E strain y: cn bw sp, *** SEQUENCING IN PROGRESS
            *** 92 unordered pieces.
AC008340 125681 bp DNA HTG 06-AUG-1999
VERSION    AC008340.2 GI:5701599
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 125681)
            Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

COMMENT
TITLE
JOURNAL
COMMENT

Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 125681)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Aug 4, 1999 this sequence version replaced gi:5670395.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to hdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 92 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1358: contig of 1358 bp in length
* 1359 1438: gap of unknown length
* 1439 2361: contig of 923 bp in length
* 2362 2441: gap of unknown length
* 2442 3575: contig of 1134 bp in length
* 3576 3655: gap of unknown length
* 3656 4853: contig of 1198 bp in length
* 4854 4933: gap of unknown length
* 4934 6297: contig of 1364 bp in length
* 6298 6377: gap of unknown length
* 6378 7092: contig of 715 bp in length
* 7093 7172: gap of unknown length
* 7173 8323: contig of 1151 bp in length
* 8324 8403: gap of unknown length
* 8404 9297: contig of 894 bp in length
* 9298 9377: gap of unknown length
* 9378 10148: contig of 771 bp in length
* 10149 10228: gap of unknown length
* 10229 11188: contig of 960 bp in length
* 11189 11268: gap of unknown length
* 11269 12360: contig of 1092 bp in length
* 12361 12440: gap of unknown length
* 12441 13007: contig of 567 bp in length
* 13008 13087: gap of unknown length
* 13088 14546: contig of 1459 bp in length
* 14547 14625: gap of unknown length
* 14626 15739: contig of 1133 bp in length
* 15740 15839: gap of unknown length
* 15840 17212: contig of 1373 bp in length
* 17213 17292: gap of unknown length
* 17293 17943: contig of 651 bp in length
* 17944 18023: gap of unknown length
* 18024 18764: contig of 741 bp in length
* 18765 18844: gap of unknown length
* 18845 19524: contig of 680 bp in length
* 19525 20464: gap of unknown length
* 20465 20544: gap of unknown length
* 20545 21397: contig of 853 bp in length
* 21398 21477: gap of unknown length
* 21478 22520: contig of 1043 bp in length
* 22521 22600: gap of unknown length
* 22601 23422: contig of 822 bp in length
* 23423 23502: gap of unknown length

```


REFERENCE 1 (bases 1 to 59255)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-22F10
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 59255)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barina, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collumore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lewczky, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,
McEwan, P., McGuirk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wymen, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L4108
Center clone name: 22_F_10

* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 754: contig of 754 bp in length
* gap of unknown length
* 755 1539: contig of 785 bp in length
* gap of unknown length
* 1540 2284: contig of 745 bp in length
* gap of unknown length
* 2285 3047: contig of 763 bp in length
* gap of unknown length
* 3048 3841: contig of 794 bp in length
* gap of unknown length
* 3842 4639: contig of 798 bp in length
* gap of unknown length
* 4640 5394: contig of 755 bp in length
* gap of unknown length
* 5395 6157: contig of 763 bp in length
* gap of unknown length
* 6158 6908: contig of 751 bp in length
* gap of unknown length
* 6909 7692: contig of 784 bp in length
* gap of unknown length
* 7693 8484: contig of 792 bp in length
* gap of unknown length
* 8485 9262: contig of 778 bp in length
* gap of unknown length
* 9263 10032: contig of 770 bp in length
* gap of unknown length
* 10033 10816: contig of 784 bp in length
* gap of unknown length

* 10817 11601: contig of 785 bp in length
* gap of unknown length
* 11602 12376: contig of 775 bp in length
* gap of unknown length
* 12377 13153: contig of 777 bp in length
* gap of unknown length
* 13154 13939: contig of 786 bp in length
* gap of unknown length
* 13940 14734: contig of 795 bp in length
* gap of unknown length
* 14735 15510: contig of 776 bp in length
* gap of unknown length
* 15511 16284: contig of 774 bp in length
* gap of unknown length
* 16285 17070: contig of 786 bp in length
* gap of unknown length
* 17071 17843: contig of 773 bp in length
* gap of unknown length
* 17844 18622: contig of 779 bp in length
* gap of unknown length
* 18623 19407: contig of 785 bp in length
* gap of unknown length
* 19408 20209: contig of 802 bp in length
* gap of unknown length
* 20210 20980: contig of 771 bp in length
* gap of unknown length
* 20981 21754: contig of 774 bp in length
* gap of unknown length
* 21755 22514: contig of 760 bp in length
* gap of unknown length
* 22515 23302: contig of 788 bp in length
* gap of unknown length
* 23303 24106: contig of 804 bp in length
* gap of unknown length
* 24107 24874: contig of 768 bp in length
* gap of unknown length
* 24875 25638: contig of 764 bp in length
* gap of unknown length
* 25639 26379: contig of 741 bp in length
* gap of unknown length
* 26380 27171: contig of 792 bp in length
* gap of unknown length
* 27172 27954: contig of 783 bp in length
* gap of unknown length
* 27955 28728: contig of 774 bp in length
* gap of unknown length
* 28729 29507: contig of 779 bp in length
* gap of unknown length
* 29508 30299: contig of 792 bp in length
* gap of unknown length
* 30300 31100: contig of 801 bp in length
* gap of unknown length
* 31101 31880: contig of 780 bp in length
* gap of unknown length
* 31881 32636: contig of 756 bp in length
* gap of unknown length
* 32637 33420: contig of 784 bp in length
* gap of unknown length
* 33421 34184: contig of 764 bp in length
* gap of unknown length
* 34185 34973: contig of 789 bp in length
* gap of unknown length
* 34974 35763: contig of 790 bp in length
* gap of unknown length
* 35764 36547: contig of 784 bp in length
* gap of unknown length
* 36548 37311: contig of 764 bp in length
* gap of unknown length
* 37312 38101: contig of 790 bp in length
* gap of unknown length
* 38102 38894: contig of 793 bp in length
* gap of unknown length
* 38895 39670: contig of 776 bp in length
* gap of unknown length

```

*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      39671      40443: contig of 773 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      40444      41210: contig of 767 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      41211      41988: contig of 778 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      41989      42772: contig of 784 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      42773      43548: contig of 776 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      43549      44334: contig of 786 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      44335      45138: contig of 804 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      45139      45923: contig of 785 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      45924      46706: contig of 783 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      46707      47484: contig of 778 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      47485      48254: contig of 770 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      48255      49016: contig of 762 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      49017      49810: contig of 794 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      49811      50603: contig of 793 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      50604      51376: contig of 773 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      51377      52156: contig of 780 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      52157      52936: contig of 780 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      52937      53728: contig of 792 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      53729      54498: contig of 770 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      54499      55296: contig of 798 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      55297      56090: contig of 794 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      56091      56878: contig of 788 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
Query Match      3.9%; Score 117.6; DB 44; Length 59255;
Best Local Similarity 96.8%; Pred. NO. 5.4e-22;
Matches 120; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 617 agacatctgtgtggagactgaactgtgtgactgatacgacgagacacatgtgtt 676
Db 33939 AGACATCTCTGTGGAACTGAACTGGTGGACTGGATGATGACGACACCATGTGTT 33998
QY 677 cactccggaactcaagctgttcgcatgtgcaagtccctgttaagaagatgtgtctcaac 736
Db 33999 CACTCCCGGACTCAAGCTGTGGCATGTGGCAATGCGCAATCCTGTTAAGAAATGGTCTTCAAC 34058
QY 737 caag 740
Db 34059 CACG 34062

```

```

RESULT 14
AC005285      209071 bp      DNA      INV      15-JUL-1998
LOCUS      Drosophila melanogaster DNA sequence (Pls DS00121 (D128), DS05470
DEFINITION      (D270), and DS00108 (D120)), complete sequence.
ACCESSION      AC005285 AC004272 AC002637 AC002638 AC003139 AC003717 AC003718
AC003140 AC003141 AC003719 AC002639 AC003142 AC003720 AC003143
AC004568 AC004269 AC002586 AC002587 AC003130 AC002588 AC002589
AC002590 AC003131 AC003709
AC005285.1      GI:3320127

```

```

KEYWORDS      HTG.
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clones DS00121
(D128), DS05470 (D270), and DS00108 (D120)) DNM.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 209071)
AUTHORS      Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Agbayani,A., Arcalata,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Kaira,K., Kearney,L.,
Kim,S.H., Lee,B., Lomolan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
Sequencing of Drosophila chromosome 2L, region 26C1-26D2
Unpublished (1997)
TITLE      2 (bases 1 to 209071)
JOURNAL      Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
AUTHORS      Svirskas,R.R., Harris,N.L., Agbayani,A., Arcalata,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Kaira,K., Kearney,L.,
Kim,S.H., Lee,B., Lomolan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
Direct Submission
Submitted (15-JUL-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
COMMENT      For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to drosophila@hgsc.lbl.gov.
Library locations: 25-2, 137-37, 12-2.
FEATURES
source      1..209071
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/map="26C1-26D2"
/clone="Pls DS00121 (D128), DS05470 (D270), and DS00108
(D120)"
/note="DS00121 (d128) extends from bp 1 to bp 84231,
DS05470 (d270) extends from bp 64,751 to bp 139,027 and
DS00108 (d120) sequenced as a bridge, extends from bp
137,172 to bp 209,071."
BASE COUNT      59585 a 45615 c 45383 g 58488 t
ORIGIN
Query Match      3.4%; Score 102.6; DB 35; Length 209071;
Best Local Similarity 53.7%; Pred. NO. 1.2e-17;
Matches 241; Conservative 0; Mismatches 199; Indels 9; Gaps 1;
QY 2392 gtttggatcgtcaactgaactcgtctgttctcagctcagaacgctgtcagctatt 2451
Db 124596 GTTTGGGTTTGATGATGATTTGGCCAGACAAACATTGTGGCCGCCGATCAAAATCGT 124655
QY 2452 aaaaaattatataagatagacagccacgtatgaagatlaaaaaatctgaattcttlt 2511
Db 124656 CAAGCAGTTTCATCAAAATTTGGCGGCACATGCAAGAGTGTCCTTAACCTTAATCATGT 124715
QY 2512 tgcacatcgtcatcagactaagaatcgtctgtgacgcgcttggcactaagctggagaa 2571
Db 124716 TGCATATCGTCTCGGGGCTTGGACACGACGACTGTTCGCGCATGCGTCAAACTGGAGAA 124775
QY 2572 actgcagaagcagttcaagaagttctatgcgagagtttgaagttaagagaccctcaag 2631

```

Db 124776 GCTGCCCTCCAAATATACAGCATGTTTCAACGACCTGCAGATWCGATGATCCCTCGCG 124835
 QY 2632 gaaccacagagcctaagcctgaagcct-----agctaagctggaacccctccat 2682
 Db 124836 CAACATGACCAAGTATCGGCACTAGTCCGCCGAACTACTGCGCCACCCCATCAT 124895
 QY 2683 ccccttcagccttccttcattaaagataagcattactactagaggggaacaagcgt 2742
 Db 124896 CCCGTTCTATCCGATGCTGAAGAGATCTACCTTTATTACCTGGGCAATGATACGAG 124955
 QY 2743 catgacaactagtaactttgaaataatgcgcattgattgaaatagcagaagcgt 2802
 Db 124956 AGTTATGAGCCTATACACTTTGAGAAAGCTGGCAATGCTGGCCAAAGAGATACGCTGCT 125015
 QY 2803 gagatactacagagcaacccctcaac 2831
 Db 125016 CACGACATGTGCGATTCGCCATACGATC 125044

RESULT 15
 AB002311 6568 bp mRNA PRI 13-FEB-1999
 LOCUS Human mRNA for KIAA0313 gene, complete cds.
 DEFINITION AB002311
 ACCESSION AB002311.1 GI:2224566
 VERSION KIAA0313.
 KEYWORDS Homo sapiens male brain cDNA to mRNA, clone_1lb:pluescriptII SK plus clone:HG0186.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 6568)
 Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M., Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
 Title Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3 Yana, Kizazazu, Chiba 292, Japan (E-mail:cdna@kazusa.or.jp, URL:http://www.kazusa.or.jp, Tel:+81-438-52-3950, Fax:+81-438-52-3931)
 2 (sites)
 Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
 Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro
 JOURNAL DNA Res. 4 (2), 141-150 (1997)
 MEDLINE 97349984
 FEATURES
 source Location/Qualifiers
 1..6568
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HG0186"
 /clone_lib="pluescriptII SK plus"
 /sex="male"
 /tissue_type="brain"
 /gene="KIAA0313"
 63..4562
 /gene="KIAA0313"
 63..4562
 /codon_start=1
 /protein_id="BAA20772.1"
 /db_xref="GI:2224567"
 /translation="MKPLAIPANHGVMQOEKSLPADFTKLHLTDSLHPQVTHVSS
 HSGSITSDSGSSSDIYOATESPAGMDLSGLPETVAVDEDDDEDIRASDPLK
 SRDIYDLERKPIDRDTDDIIELEFMHLPAFANMTMSVRRLECAVMFAVVERAC
 TYIWDGEELDSMSYIINGSVYTPYDPKAEILCKGNSFGSPYDRKRYMGVARTKY
 DDCQFVIAQDYCRILNQVERNMOKVEEELVAVKSHRELDRTRKGIIVKGIS
 EULTMHLEHSHVDPPTFIEDFLVYTRFLSSPMVEGKLLNLNLAACAARRLMTLTKPS
 LMVNHFNDFEGDPAFTFIEFENNELREKMGHLRLNLNLAACAARRLMTLTKPS
 REAPLPFLLIGSEKGFIFVDSVDSGSKATEAGLKRGLLELVNGQFENIOLSKAM
 ELRNNTHLSTIVKTNLFPVKELLRLRSEKRNAPHLPKIGDIKKARSYISPLAVD

VEOYIGLEKYNKSKANTVGRNRKLLKIDKTRISILPOKPVNDIGIOSODSIYGL
 ROKKHIPALPVSGTSSSNPDLOSRRILDSAPBDLPDYLRFKPDQOSRYTMI
 SKOTTAKEVYIOAIRERAVATATPDQISLSTVSPSVIKORLIPQSLKARIDLS
 GRYYLKNMGETETLDSDEQELLRESQISLQSLSTVEATQLSMNFELFRITEPE
 YIDLEFKRSKTSVCANLREPEVINOTEVASILLERINOLKRMKTIHFKIALAC
 RECKNFMSFAIISGLNLAPEVALRTTWELIPYKRELFODLLODLDPDRMAKYRY
 LNSONLOPPIIPLPEYIKDITFLHEGNSKVSIVNQTGHNKRELFIAKIEHVMSVN
 MDPALMRTKKKRWRSLSGSOSTNATYVDVOTGCHKRKRVSFLNAKUYEDAO
 MARKVOYLSNLEEMDEBSLQTLSDCEPATITLERNPQDKPVKSETSPVAPRAGS
 QOKAOSLPQOQPPPAHKTINQILOVPVAVSLYPSRKVPYKDPFGINSPOLKIL
 SLSEBSLERHKKQAEEDTISNASSQSLSPSTSPSRKGYTLAPGIVNSPOLKIL
 EISRSISIVNSSEFVPSLHDERORHSVIVETNLGGRERRTMTLEPDYSLGS
 YAPWSEGRGLYATATVYISPSSTELSDODGRASLILOAASGRSMTSCSSGNDNOT
 IOHORSMTPEHHTHDYSGDPAAGLMASSSHNDQIMEFSHSTKYRQMSRSLQIA
 QSRASMASSTGVWGEDSDGTIKRKGADVSIEAESLSVSTTEKTPVMPRAHI
 AVASSTTKGLIARKEGRREPPPTPEYIGIIPITDPEEGSHPARPPPYNVALOHR
 MVARSPTDGPSSVQOPPHGHTSSRPVKNPKMQRNPRPRLAPYOSOGFSREDEDE
 OVSAY"

BASE COUNT 1974 a 1400 c 1463 g 1731 t
 ORIGIN
 Query Match 3.3%; Score 99.2; DB 9; Length 6568;
 Best Local Similarity 51.1%; Pred. No. 5.9e-17;
 Matches 289; Conservative 0; Mismatches 268; Indels 9; Gaps 2;
 QY 2228 ctgatgagctccaaagattgagcataccagaatgacaattatgatgtggaactctaac 2287
 Db 2205 CAGCTCACGACTGTGGAGAGTTGGAAGTTCACACGCTCTATGCGAAATTTWGAACCTCTTCGC 2264
 QY 2288 tgcgagctgctgagcagctaatcatatcacacatttggaagcgaatatttaaaagac 2347
 Db 2265 AACATTGACCTACTGAAATATATGATGATTTATTT--AAACTCAGATCAAAACACAGC 2331
 QY 2348 aacgaacacttgattgttctctgagagagattgaatgaattcagtttggctgctact 2407
 Db 2332 TGTGCAACCTGGAAGAGATTGGAAGAGTATTACCAAGAAATTTTGGTGATCT 2381
 QY 2408 gagatgctccttctctcagctcagaacagcgcttctcagctattataaaatttaag 2467
 Db 2382 GAAATTCACAGAAACCAACACGCTGAGAGATGGAAGATCAATTAAGCATTTTCATCAG 2441
 QY 2468 atagagcccccctgtaagagatataaaatctgattcccttttggccatcgatagga 2527
 Db 2442 ATAGCACTGCTGAGGAAATGCAAGATTTTAACTCAATGTTTCAATCATCATGAGGC 2501
 QY 2528 ctaagtaacattgctgtagcgccttgcaactaagctgaggaacgcaagcaagctc 2587
 Db 2502 CTAAACCTGGCAGCAGTGGCAAGACTGCGAACGACTGGGAGAACTTCCCAATTAATAC 2561
 QY 2588 aagaagttctatgctgagatttgaagtttaatggaaccttaagagcaacagggcctac 2647
 Db 2562 GAAAAACTATTTCAAGATCTCCAGACCTGTTGATCTCTCCAGAAACATGGCAAAATAT 2621
 QY 2648 ag-----gtgacaagtagcctgaagcctctctcatcccttcacagccttgcctc 2701
 Db 2622 CGTATATGTTCAATAGTCAAAATCTACAACTCCATTAATCCCTTATTCCTCCAGTATC 2681
 QY 2702 attaaagatagacattactacatgaggggaacaagcgcgttacttggaacatctagtaac 2761
 Db 2682 AAAAAGATCTCACCTTCTCAAGAGAAATATCATCAAAAGTAGACGGGCTGTCAT 2741
 QY 2762 ttgaaaaaatgcgcatgattgcaaa 2787
 Db 2742 TTTGAAGAGTAAAGATGATTGCAAA 2767

Search completed: April 22, 2000, 17:50:32
 Job time: 32288 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 16:15:23 ; Search time 351.53 Seconds
(without alignments)
1026.159 Million cell updates/sec

Title: US-09-422-999-17
Perfect score: 3013
Sequence: 1 gatccagcgagatgtgat.....ccacatttcaaatgcc 3013

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5C.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5D.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/6.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/PTC059.COMB.seq:*
7: /cgn2_6/ptodata/2/1na/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.8	1.5	7218	1	Sequence 14, Appl
2	35.4	1.2	2652	1	Sequence 1, Appl
3	35.2	1.2	5183	4	Sequence 7, Appl
4	34.6	1.1	2128	3	Sequence 16, Appl
5	34.6	1.1	2338	2	Sequence 1, Appl
6	34.6	1.1	2338	4	Sequence 1, Appl
7	34	1.1	4368	1	Sequence 17, Appl
8	34	1.1	4446	1	Sequence 6, Appl
9	33.4	1.1	4937	1	Sequence 3, Appl
10	33.4	1.1	4937	1	Sequence 3, Appl
11	33.4	1.1	4937	3	Sequence 3, Appl
12	33.4	1.1	4937	3	Sequence 3, Appl
13	33.4	1.1	4937	3	Sequence 3, Appl
14	33.4	1.1	4937	4	Sequence 3, Appl
15	33.4	1.1	4937	4	Sequence 3, Appl
16	33.4	1.1	9323	1	Sequence 6, Appl
17	33.4	1.1	9323	3	Sequence 6, Appl
18	33.4	1.1	9323	3	Sequence 6, Appl
19	33.4	1.1	9323	4	Sequence 6, Appl
20	33.4	1.1	9323	4	Sequence 6, Appl
21	33	1.1	5225348	2	Sequence 6, Appl
22	32.8	1.1	1302	3	Sequence 1, Appl
23	32.8	1.1	1302	4	Sequence 1, Appl
24	32.6	1.1	2277	2	Sequence 5, Appl
25	32.6	1.1	2277	2	Sequence 5, Appl
26	32.6	1.1	2277	4	Sequence 5, Appl
27	32.4	1.1	175	1	Sequence 58, Appl

28	32.4	1.1	175	2	US-08-474-633A-76	Sequence 76, Appl
29	32.4	1.1	175	6	PCT-US92-06412-58	Sequence 58, Appl
30	32.2	1.1	1401	6	PCT-US92-06965A-32	Sequence 32, Appl
31	32	1.1	160	1	US-08-182-175A-32	Sequence 32, Appl
32	32	1.1	160	1	US-08-182-175A-34	Sequence 34, Appl
33	32	1.1	160	2	US-08-474-633A-31	Sequence 29, Appl
34	32	1.1	160	2	US-08-474-633A-31	Sequence 31, Appl
35	32	1.1	160	6	PCT-US92-06412-32	Sequence 32, Appl
36	32	1.1	160	6	PCT-US92-06412-34	Sequence 34, Appl
37	32	1.1	187	1	US-08-182-175A-90	Sequence 90, Appl
38	32	1.1	187	1	US-08-474-633A-78	Sequence 78, Appl
39	32	1.1	187	6	PCT-US92-06412-90	Sequence 90, Appl
40	31.8	1.1	660	2	US-08-307-279A-11	Sequence 11, Appl
41	31.8	1.1	2277	2	US-08-676-974-2	Sequence 2, Appl
42	31.8	1.1	2277	2	US-08-676-974-2	Sequence 2, Appl
43	31.8	1.1	2277	4	US-09-098-487-2	Sequence 2, Appl
44	31.8	1.1	2356	1	US-08-105-483-222	Sequence 222, App
45	31.8	1.1	2356	1	US-08-220-151-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

```

Query Match      1.5%; Score 44.8; DB 1; Length 7218;
Best Local Similarity 7.9%; Pred. No. 0.0015;
Matches 34; Conservative 208; Mismatches 190; Indels 0; Gaps 0;

Qy 1704 atgaagctatgagccttccttgaggaaggttttattgtatcgtatcagatgataccgga 1763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1480 ATTACCTATCTATGCAAGTACTTAAGAGATGAAGAATTGTGACRRRRRRRRRRRRRRR 1421

Qy 1764 tgatgctgcctcaagagcaactgagcagatctggaagaatgctcaagaatcctcag 1823
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1420 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1361

Qy 1824 aagatgcaagagccacacacaaagaagcaagctcttcttcaacagatcacaagcg 1883
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1360 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1301

Qy 1884 atgagagagccagagaagcgacgtatcgcgcgtctgtatgaagtcgttgaagctc 1943
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1300 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1241

Qy 1944 atgcatgacccacacacacacacacatcggtcggtcgacatcctcggtgaagga 2003
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1240 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1181

Qy 2004 tcatcagtcagctgagcagacagctcggtcggtcggtcggtatcatagatcaagatga 2063
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1180 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1121

Qy 2064 gtccgagagagaaagtgctcaaacatgatatgcttgcgttcttaagagctca 2123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1120 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1061

Qy 2124 ccatatgagac 2135
    | ||| |
Db 1060 AGCTCCCTCGAC 1049

RESULT 2
US-08-318-831-1
; Sequence 1, Application US/08318831
; Patent No. 5656595
; GENERAL INFORMATION:
; APPLICANT: Schweighoffer, Fabien
; APPLICANT: Tocque, Bruno
; TITLE OF INVENTION: PEPTIDES HAVING A GDP EXCHANGE FACTOR
; TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SAID PEPTIDES.
; TITLE OF INVENTION: PREPARATION AND UTILIZATION
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: word 5.1 (Epo PatentIn)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,831
; FILING DATE: 19 October 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR92/04827
; FILING DATE: 21-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K. P-38,619
; REGISTRATION NUMBER: P-38,619
; REFERENCE/DOCKET NUMBER: ST92033-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 454-3839

```

```

; TELEFAX: (610) 454-3808
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2652 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2445
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 445..2445 (SEQ ID NO 3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 976..2445 (SEQ ID NO 4)
; US-08-318-831-1

Query Match      1.2%; Score 35.4; DB 1; Length 2652;
Best Local Similarity 42.5%; Pred. No. 0.72;
Matches 247; Conservative 0; Mismatches 331; Indels 3; Gaps 1;

Qy 2220 ctttgaactgtagcctcccaagatttagcataccagatgacatttattgttggaaac 2279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1721 CTTTGAAGAAACCATGACCCCTTGAGATGCGGAGACGCTGACCTGTATGATCACTCG 1780

Qy 2280 tctcaactgcygcatgagctgagctaatctatcatcacacatttgaagacataattta 2339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1781 TCTTCAAGAAAGATTCTTTAGAGAGATTCTTGGACAGAGATGATGAAGTCAAGAAAGA 1840

Qy 2340 aaaaagaccacgaacattgattgttctctgaagagatttaagaaattcaatttggg 2399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1841 ATGAAGAGACCCCTTATATCATGAAACACATAGCATTAAGCATTAAGCATTAAGTATGA 1900

Qy 2400 tctcaactgagatctgcttcttctcagctcagcaagctgttcaagctatttaaaaaat 2459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1901 TTGCTTCAGAAATCATCCCAATGAGAGATCAACCCAGGTGAGCCCATGAGAGT 1960

Qy 2460 ttaataagtagagccacatgaaagagatataaatactgatactcttttgcacag 2519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1961 GGGTGGCCGATGATGATATGCGGTGCTCCACACATACATGCGGTACTGAGATCA 2020

Qy 2520 tcatggaactgaatcatctgctgtgagccgttggcactaagctggagaaactggcaa 2579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2021 CCTGCTCATGAACCGCATGCAATCTCCGGCTCAAAAAGACGTGCTCAAAAGTCTCTA 2080

Qy 2580 gcaagttcaagaagttctatgagagtttgaagttaatggaaccttcaagaaccaca 2639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2081 AGCAGACTAAAGCTTGATGATTAAGCTCAAAAAGCTTGATGATGAGGCAATTTA 2140

Qy 2640 gggccctacagctgacagtagctgaagctcctcctcctcctcctcctcctcctg 2699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2141 AGAATCTCAGAGAGCTTTGAAAAATTGACCCACCTGTTCCCTTACCTGGGAGATG 2200

Qy 2700 tcaataagatatgacatttactcatgagagggagaaacaaagacgttca--ttgacaactcag 2756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2201 ACCTCACACGACCTGGCTTCATCGAGAGGAGACGGCCCAATTACGGAAGACGGCCCTGG 2260

Qy 2757 taactttgaaaaaatgagcatgattgcaatacagggcaga 2797
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2261 TCAACTTCTCAAGATGAGATATATCCCATATTATTCGA 2301

RESULT 3
US-08-870-518-7
; Sequence 7, Application US/08870518
; Patent No. 5925566
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.

```


APPLICANT: Galcheva-Gargova, Zoya
TITLE OF INVENTION: NON-ACTIVATED RECEPTOR COMPLEX
TITLE OF INVENTION: PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870.518
FILING DATE: 06-JUN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,219
FILING DATE: 06-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/102001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 5183 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-870-518-7

Query Match 1.28; Score 35.2; DB 4; Length 5183;
Best Local Similarity 54.7%; Pred. No. 1.4;
Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2578 aagcaagttcaagaagttcgtcgaggttgaagtttaagacccttcaagagacaa 2637
DB 1970 AAGACTGGTGGTGAATTCACCCAAAGTCGAAAATTTAAAGTCATGATGCC 2029
QY 2638 cagggcctacagcgtacagtaagctgaacctctctcatcccttcagctt 2697
DB 2030 GAGGACTATCCGCTGATATCTCAATTCGAAACCGCATCTCTTAAATTCCTGAACCT 2089
QY 2698 gctcatta 2705
DB 2090 GGACTTGA 2097

RESULT 4
US-08-371-377-16
; Sequence 16, Application US/08371377
; Patent No. 5851764
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Shen, Ruogian
; TITLE OF INVENTION: DEVELOPMENT OF DNA PROBES AND
; TITLE OF INVENTION: IMMUNOLOGICAL REAGENTS SPECIFIC FOR CELL SURFACE-EXPRESSED
; MOLECULES AND TRANSFORMATION-ASSOCIATED GENES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/371.377
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/37590-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ. ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2128 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 621..1817
US-08-371-377-16

Query Match 1.1%; Score 34.6; DB 3; Length 2128;
Best Local Similarity 54.3%; Pred. No. 1.1;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 100 gttattatgagatctggaagaaggaataacattatttcgccaggtgatatagaacaa 159
DB 1573 GTTCTGTTAAAGCTGAGAGATGGCCCTTAATCTTGAAGTCTGGATCTCCATTTG 1632
QY 160 ctgtatctgcctgcagaggtcttggatgttaagatctgagaccagtcacaa 219
DB 1633 TTGATATGTTCTCTGCGAAGCCATGCTGTGTGAGAGCTTCTCAGACTATCCACTTTGG 1692
QY 220 gatcgtctg 228
DB 1693 GCTGCTTTG 1701

RESULT 5
US-08-425-069-1
; Sequence 1, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; MOLECULES AND TRANSFORMATION-ASSOCIATED GENES
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25


```
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nishiohata Residence 1-107
;; STREET: 5214, Nishiohata-machi
;; CITY: Niigata-shi
;; STATE: Niigata-ken
;; COUNTRY: JAPAN
;; ZIP: 951
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS v.5
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/026.138E
;; FILING DATE: 26-FEB-1993
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 39563/1992
;; FILING DATE: 26-FEB-1992
;; APPLICATION NUMBER: JP 173155/1992
;; FILING DATE: 30-JUN-1992
;; APPLICATION NUMBER: JP 215017/1992
;; FILING DATE: 12-AUG-1992
;; APPLICATION NUMBER: JP 303878/1992
;; FILING DATE: 13-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hamburg, C. Bruce
;; REGISTRATION NUMBER: 22,389
;; REFERENCE/DOCKET NUMBER: F-4551
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 986-2340
;; TELEFAX: (212) 953-7733
;; INFORMATION FOR SEQ ID NO: 17:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4368 nucleic acids
;; TYPE: nucleic acid
;; STRANDEDNESS: double strand
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; ORIGINAL SOURCE:
;; ORGANISM: mouse
;; TISSUE TYPE: brain
;; PUBLICATION INFORMATION:
;; AUTHORS: Masayoshi MISHINA
;; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
;; RELEVANT RESIDUES IN SEQ ID NO: 17: FROM 1 TO 4368
;;
US-08-026-138E-17

Query Match 1.1%; Score 34; DB 1; Length 4368;
Best Local Similarity 54.9%; Pred. No. 2.9;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1072 gtttaaccagggggaagaagctacctcctggtacattattcttaaaagaagtcagtgaatgt 1131
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2046 GTTCAACCAAGGGGTGTAGATGAGCCCTTGCTGCCCTGAAACAGGAAACTTGATGC 2105

QY 1132 agtcattacggcgaaggtgtgtctgcaacctgcatgaagaagatgacttcggcaagt 1191
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2106 ATTCACTACGATGACGACCGGTGCTCAACTACATGCTGGAGAGACGAGGCTGCAAGCT 2165

QY 1192 ag 1193
DB 2166 GG 2167

RESULT 8
US-08-026-138E-6
; Sequence 6, Application US/08026138E
; Patent No. 5502166
; GENERAL INFORMATION:
; APPLICANT: Masayoshi MISHINA
; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
; NUMBER OF SEQUENCES: 19
```

```
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nishiohata Residence 1-107
;; STREET: 5214, Nishiohata-machi
;; CITY: Niigata-shi
;; STATE: Niigata-ken
;; COUNTRY: JAPAN
;; ZIP: 951
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS v.5
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/026.138E
;; FILING DATE: 26-FEB-1993
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 39563/1992
;; FILING DATE: 26-FEB-1992
;; APPLICATION NUMBER: JP 173155/1992
;; FILING DATE: 30-JUN-1992
;; APPLICATION NUMBER: JP 215017/1992
;; FILING DATE: 12-AUG-1992
;; APPLICATION NUMBER: JP 303878/1992
;; FILING DATE: 13-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hamburg, C. Bruce
;; REGISTRATION NUMBER: 22,389
;; REFERENCE/DOCKET NUMBER: F-4551
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 986-2340
;; TELEFAX: (212) 953-7733
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4446 nucleic acids
;; TYPE: nucleic acid
;; STRANDEDNESS: double strand
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; ORIGINAL SOURCE:
;; ORGANISM: mouse
;; TISSUE TYPE: brain
;; PUBLICATION INFORMATION:
;; AUTHORS: Masayoshi MISHINA
;; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
;; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 4446
;;
US-08-026-138E-6

Query Match 1.1%; Score 34; DB 1; Length 4446;
Best Local Similarity 54.9%; Pred. No. 2.9;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1072 gtttaaccagggggaagaagctacctcctggtacattattcttaaaagaagtcagtgaatgt 1131
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2124 GTTCAACCAAGGGGTGTAGATGATGCTTGCTGCCCTGAAACAGGAAACTTGATGC 2183

QY 1132 agtcattacggcgaaggtgtgtctgcaacctgcatgaagaagatgacttcggcaagt 1191
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2184 ATTCACTACGATGACGACCGGTGCTCAACTACATGCTGGAGAGACGAGGCTGCAAGCT 2243

QY 1192 ag 1193
DB 2244 GG 2245

RESULT 9
US-08-038-682-3
; Sequence 3, Application US/08038682
; Patent No. 5549697
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
```

```

; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038.682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; US-08-038-682-3

```

```

Query Match      1.1%; Score 33.4; DB 1; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

```

Qy 2239 caaagattagcataccagatgacaattatgatgggaactcttcaactgcygcatga 2298
    ||||| ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2409 CAAAGGCTTAAACAACAGTATAGAGCTCTGCAGCGGTGAATTTTAAACGCGTAATG 2468
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2299 gctggagctaatctacacacatttggaagcagcacaatttaaaagaccagcaact 2358
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2469 CAACATGTCATTCATTCCTCAAGAAGAGCGAAGTATTCAATTTAAATTAACCAACGA 2528
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2359 ggaattgtcctcgtgagagattaatgaattcaagtttggc 2401
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2529 GAACATGAACACAGCAAACTTTACCAATTCGGTTTTTAACC 2571
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
US-08-302-832-3
; Sequence 3, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; US-08-302-832-3

```

```

Query Match      1.1%; Score 33.4; DB 1; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

```

Qy 2239 caaagattagcataccagatgacaattatgatgggaactcttcaactgcygcatga 2298
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2409 CAAAGGCTTAAACAACAGTATAGAGCTCTGCAGCGGTGAATTTTAAACGCGTAATG 2468
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2299 gctggagctaatctacacacatttggaagcagcacaatttaaaagaccagcaact 2358
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2469 CAACATGTCATTCATTCCTCAAGAAGAGCGAAGTATTCAATTTAAATTAACCAACGA 2528
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2359 ggaattgtcctcgtgagagattaatgaattcaagtttggc 2401
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2529 GAACATGAACACAGCAAACTTTACCAATTCGGTTTTTAACC 2571
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-08-530-198-3
; Sequence 3, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424

```

```
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4937 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-530-198-3
```

```
Query Match 1.1%; Score 33.4; DB 3; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
QY 2239 caaagattagcatcacagatgacattatgattggaaactctcaactcgtcatga 2298
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2409 CAAGCGCTTAACAACACAGTATAGAGCTCTGCAGGGTGATTTTACGCGCTTAATGG 2468
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2239 gctgagctaatctatcacacatttgaaggcataatttaaaagaccacagcaactt 2358
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2469 CAACATGTCATTCATTCACAAAGAGGCGAAGTTAATTTCAAATTAAACCAACGA 2528
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2359 gatttctcctgaggagatttaagtaattcagtttgggtc 2401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2529 GAACATGAACACAGCAACCTTACCAATTCGGTTTAAAGCC 2571
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-08-469-880-3
Sequence 3, Application US/08469880
Patent No. 587673
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 587673-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4937 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-3
```

```
Query Match 1.1%; Score 33.4; DB 3; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
QY 2239 caaagattagcatcacagatgacattatgattggaaactctcaactcgtcatga 2298
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2409 CAAGCGCTTAACAACACAGTATAGAGCTCTGCAGGGTGATTTTACGCGCTTAATGG 2468
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2239 gctgagctaatctatcacacatttgaaggcataatttaaaagaccacagcaactt 2358
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2469 CAACATGTCATTCATTCACAAAGAGGCGAAGTTAATTTCAAATTAAACCAACGA 2528
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2359 gatttctcctgaggagatttaagtaattcagtttgggtc 2401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2529 GAACATGAACACAGCAACCTTACCAATTCGGTTTAAAGCC 2571
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 13
US-08-728-470-3
Sequence 3, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 1038-633
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 3:
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 16:07:45 ; Search time 625.03 Seconds
(without alignments)
1206.068 Million cell updates/sec

Title: US-09-422-999-17
Perfect score: 3013
Sequence: 1 gatccagcgaagatgtgat.....ccacatttcaaaatgcc 3013

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.6	1.7	335	1 V90407	EST clone DL605. N
2	43.6	1.4	4062	1 T13347	C3G protein gene.
3	36.4	1.2	110000	1 V21209_12	Continuation (13 o
4	35.4	1.2	2388	1 V84554	Human secreted pro
5	35.4	1.2	2652	1 051233	Sequence encoding
6	35.4	1.2	9601	1 N80860	Sequence corresp.
7	35.2	1.2	5183	1 V04063	Schizosaccharomyce
8	34.6	1.1	2128	1 T37405	Prostate tumour in
9	34.6	1.1	2338	1 Q14183	N.clavipes draglin
10	34.6	1.1	2338	1 V23249	Nephila clavipes s
11	34.6	1.1	4833	1 V65139	Ethlichia sp. W20
12	34.6	1.1	22243	1 V74475	Staphylococcus aur
13	34.4	1.1	1677	1 V06059	Human imidazoline
14	34.4	1.1	1954	1 X28368	Human imidazoline
15	34.4	1.1	3317	1 V06060	Human imidazoline
16	34.4	1.1	3318	1 X28369	Human imidazoline
17	34.4	1.1	3389	1 X28367	Human imidazoline
18	34.4	1.1	15202	1 X28371	Human imidazoline
19	34.2	1.1	110000	1 X20248_04	Continuation (5 of
20	34	1.1	4368	1 055979	NMDA receptor chan
21	34	1.1	3395	1 V05477	Glutamic acid rece
22	33.6	1.1	3395	1 V05477	Homo sapiens secre
23	33.4	1.1	1872	1 Q39220	Carrot Reverse tra
24	33.4	1.1	4937	1 Q43507	Sequence encoding
25	33.4	1.1	4937	1 Q72294	Haemophilus high m
26	33.4	1.1	4937	1 T90995	Non-tyeable Haemo
27	33.4	1.1	9323	1 Q49509	Gene cluster for h
28	33.4	1.1	9323	1 T90997	Non-tyeable Haemo
29	33.4	1.1	9502	1 Q74770	Hepatitis C virus
30	33.2	1.1	524	1 T35109	Down-regulated sen
31	33.2	1.1	524	1 T35109	L.lactis HsdS gene
32	33	1.1	581	1 X29775	Human gene signatu
33	33	1.1	788	1 V69012	DNA molecule encod
34	33	1.1	825	1 Q27006	HK4. Hepatitis C v

35	33	1.1	1753	1 Q06413	Sequence encoding
36	33	1.1	9360	1 Q02830	cDNA to HIV-2 RNA.
37	32.8	1.1	1302	1 T46127	Maga gene. maga ge
38	32.8	1.1	1302	1 T90222	Magnetospirillum s
39	32.8	1.1	1312	1 V27047	Xenopus neurogenin
40	32.6	1.1	1397	1 T84935	Human prostate pro
41	32.6	1.1	1797	1 T84931	Human prostate pro
42	32.6	1.1	2162	1 X14106	H. pylori GHPO 16
43	32.6	1.1	2277	1 V05372	Human telomerase p
44	32.6	1.1	2277	1 V13836	Homo sapiens mamma
45	32.4	1.1	175	1 Q37280	Clone 5-1 used in

ALIGNMENTS

RESULT 1	
ID V90407	Standard; cDNA; 335 BP.
AC V90407:	
DT 15-FEB-1999	(first entry)
DE EST clone DL605.	
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;	
KW tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;	
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;	
KW gene therapy; ss.	
OS Homo sapiens.	
PN W09845436-A2.	
PD 15-OCT-1998.	
PF 10-APR-1998; U06955.	
PR 10-APR-1997; US-838821.	
PA (GENEX) GENETICS INST INC.	
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,	
PI Racie JA, Spaulding V, Treacy W;	
DR WPI; 99-070077/06.	
PT New polynucleotides encoding human secreted proteins - derived from	
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,	
PT ovary, pituitary, retina and colon cDNA libraries.	
PS Claim 1; Page 530-531; 618pp; English.	
CC The present sequence represents a human expressed sequence tag (EST).	
CC The polynucleotide, which is a secreted EST, and the encoded protein	
CC are predicted to have useful biological activities which would make	
CC them suitable for treating, preventing or ameliorating medical	
CC conditions in humans and animals, although no supporting data is	
CC given. Suggested activities include nutritional activity, immune	
CC stimulating or suppressing activity, haematopoiesis regulating	
CC activity, tissue growth activity, activin/inhibin activity,	
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic	
CC activity, receptor/ligand activity, anti-inflammatory activity,	
CC cadherin/tumour invasion suppressor activity, tumour inhibition,	
CC activity. The polynucleotide may also be useful for gene therapy.	
SC Sequence 335 BP; 118 A; 53 C; 78 G; 86 T;	
SQ	
Query Match 1.7%; Score 50.6; DB 1; Length 335;	
Best Local Similarity 55.3%; Pred. No. 2e-05;	
Matches 120; Conservative 0; Mismatches 94; Indels 3; Gaps 1;	
QY 582 tgaatagatagaataaccacccaagaacatacgaacatgctgtgtggaactgac 641	
DB 79 tgaatgaagacagaaaaaagaaacgactacgttccctaaatgctttttggaagccaat 138	
QY 642 tggtagactgtagatcagcagacacatgcttaccctccggaactcaagctgttgga 701	
DB 139 ttgtcttcattgctgtt---ggaaattgagagattcacagccctgaggaagccgtgcact 195	
QY 702 tgtgcaagctcgttgaagaatggtgttccacacagctgagcagagacacattcc 761	
DB 196 tgggcaagcattatttgaataaagaaatcattcaccagttactgattgaataattca 255	
QY 762 aagacttattattcattcattcattcattcattcattcattcattcattcattcattc 798	
DB 256 aacgaaacacagatgtttattatgatattccgtatgatga 292	

```
RESULT 2
ID T13347
ID T13347 standard; cDNA to mRNA; 4062 BP.
AC T13347;
DR 26-JUL-1996 (first entry)
DE C3g protein gene.
KW C3g; ras protein guanine nucleotide exchange factor; diagnosis;
  treatment; malignant tumour; activation; ss.
OS Homo sapiens.
FH key Location/Qualifiers
FT cds 123..3356
      /*tag=
      /note= "C3g protein"
PN J07051073-A.
PD 28-FEB-1995.
PF 13-JUN-1994; 130699.
PR 11-JUN-1993; JP-140806.
PA (SHKJ) SHINGIJUTSU JIGYODAN.
DR WPI: 95-127357/17.
P-PEDB: R91597.
PT Ras protein guanine nucleotide exchange factor C3g gene - useful for
  diagnosis and treatment of malignant tumours associated with ras
  oncogene activation
PS Claim 2; Page 6; 9pp: Japanese.
CC The present sequence encodes C3g protein which is a ras protein guanine
  nucleotide exchange factor. The gene and protein (see R91597) are useful
  for diagnosis or treatment of malignant tumours associated with
  activation of the ras gene.
SQ Sequence 4062 BP; 937 A; 1194 C; 1147 G; 784 T;
```

```
Query Match 1.4%; Score 43.6; DB 1; Length 4062;
Best Local Similarity 46.2%; Pred. No. 0.012;
Matches 197; Conservative 0; Mismatches 214; Indels 15; Gaps 1;

QY 2338 taaagaaccacagaacatgattgttctcaggagagattgaatcagtttg 2397
DB 2738 taaagaaccacagaacatgattgttctcaggagagattgaatcagtttg 2397
QY 2398 ggtcgtcactgagatcgtctgttctcagtcagcaagcgtgttcagctataaaaaa 2457
DB 2798 ggtcgtcactgagatcgtctgttctcagtcagcaagcgtgttcagctataaaaaa 2457
QY 2458 atttaagaagaagcagccctctgaagagatataaattgaattcctttttgcat 2517
DB 2858 gtttcattcaagatcagaaagcagcttcggaagcgtgaatacttcaactctctggtgcat 2517
QY 2518 cgtcatggagactgaactgaatcgtgagccgttgacactgaagtggaagaactgcc 2577
DB 2918 cgtcatggagactgaactgaatcgtgagccgttgacactgaagtggaagaactgcc 2577
QY 2578 aagcaagtcagaagatctatgcgagattgaagaatttaatgaagccttcaagaacca 2637
DB 2963 gaagcagacttcagagagcgttcgagagatgcacactgactgacacgtgctcctctt 3022
QY 2638 caggagcctacagctgacagatgaactgaagcctctcctcctcctcagctt 2697
DB 3023 cggagcctacagctgacagatgaactgaagcctctcctcctcctcagctt 2697
QY 2698 gtcattaaagatgacatttactcagagagagagaagaagcagttcagcaatctagt 2757
DB 3083 gattcctgcagagactgacattgcttgcacactgaggaagaccagactgacagcggaagt 3142
QY 2758 aaactt 2763
DB 3143 GAACCTT 3148
```

```
RESULT 3
V21209_12/c
Continuation (13 of 17) of V21209 from base 1200001 (Methanococcus jannaschii circular
```

```
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976
```

```
Query Match 1.2%; Score 36.4; DB 1; Length 110000;
Best Local Similarity 49.0%; Pred. No. 14;
Matches 97; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1316 agacttaagaacatgacccaagatgcttgcgtcgtgagaaggtccacagaggaacaga 1375
DB 69595 AAACCTAATTCGAAACCAACATCGATGATTAATTTGATTAATTCGAAACCAACGAAATTA 69536
QY 1376 gcttcaatacagaagaactacagcctcagcaaaagtatactgtgactgaagaacact 1435
DB 69535 AATTAACATTAATTCGAAACCAACGAGAGTTTGTAAATGCAATTTAGATATAAGACTTAGT 69476
QY 1436 gaaaaattttagagacatttctagaacaatacgccttgaggcaacttaatgaagca 1495
DB 69475 GAAAAAATTAATTTTAAAGATGTGAAGTTGTATATCATCAACGACCTCAAAATTAACGTT 69416
QY 1496 acagatctgtttaaat 1513
DB 69415 AGAATTTCTGTTGAAT 69398
```

```
RESULT 4
V84554/c
ID V84554 standard; DNA; 2388 BP.
AC V84554;
DR 01-MAR-1999 (first entry)
DE Human secreted protein gene 144 clone HGLAM46.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
  diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
  developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
  immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
  inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
  cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
  endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN W09854963-A2.
PD 10-DEC-1998.
PF 04-JUN-1998; 0111422.
PR 18-DEC-1997; US-070923.
PR 06-JUN-1997; US-048877.
PR 06-JUN-1997; US-048881.
PR 06-JUN-1997; US-048884.
PR 06-JUN-1997; US-048893.
PR 06-JUN-1997; US-048896.
PR 06-JUN-1997; US-048899.
PR 06-JUN-1997; US-048915.
PR 06-JUN-1997; US-048949.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048972.
PR 06-JUN-1997; US-049020.
PR 06-JUN-1997; US-049375.
```

CC	encoding human secreted proteins (W8534 to W8755). The secreted protein
CC	gene sequences are deposited with the ATCC under deposit numbers ATCC
CC	97979, 97974, 97976, 97976, 97977, 209007, 209008, 209009, 209010,
CC	209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC	cells comprising recombinant vectors containing the nucleic acid
CC	sequences are used for the recombinant production of the secreted
CC	proteins. The polynucleotide and amino acid sequences are useful for
CC	useful for preventing, treating or ameliorating medical conditions e.g.
CC	by protein or gene therapy. Pathological conditions can be also diagnosed
CC	by determining the amount of the new polypeptides in a sample or by
CC	determining the presence of mutations in the new polynucleotides.
CC	Specific uses are described for each of the polynucleotides, based on
CC	which tissues they are most highly expressed in, and include developing
CC	products for the diagnosis or treatment of cancer, neurodegenerative
CC	disorders, developmental abnormalities and foetal deficiencies, blood
CC	diseases, tumours, leukemias, diseases of the immune system, autoimmune
CC	diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC	ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC	restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC	such as osteoporosis, arthritis or malignancies, diseases of testes, lung
CC	or thymus, digestive/endocrine disorders, infections and AIDS. The
CC	polypeptides are also useful for identifying their binding partners.
CC	The present sequence represents a gene encoding a human secreted protein
CC	(see descriptor line for gene number and clone identification).
SQ	Sequence 2388 BP; 703 A; 471 C; 560 G; 653 T;
Query Match	1.2%; Score 35.4; DB 1; Length 2388;
Best Local Similarity	51.6%; Pred. No. 2.6;
Matches 81; Conservative	0; Mismatches 76; Indels 0; Gaps 0
OY	2503 ttcccttttgcacatgcatggactagtaacgttcgttgagcgcgttgacatac 2562
Db	1874 TGCCACATTTTACACACACCCTGCGCAAGTAATTATTTTTTTTCATGTAGAACAATAA 1915
OY	2503 gtgggagaacaactccaagcaagtccaagaagttcctaibcgagagttgaagtttaatgga 2622
Db	1814 TATGCGCAGCAAAAGTAAGTAGTTAGGAAGTGCTGTGGTGGGGAATGTGTAAAA 1755
OY	2623 cccttcaaggaccacacagggccctacaggtcacagta 2659
Db	1754 TCCTTTAAAGGATTAAGGGAATTAAGTGGCTTTTATA 1718
RESULT	5
ID	051233 standard; cDNA; 2652 BP.
AC	051233;
DT	11-MAY-1994 (first entry)
DE	Sequence encoding peptides which modulate GDP exchange.
KW	GDP; exchange; GRF; GDP exchange factor; antisense; diagnosis;
RW	detection; cancer; ras; oncogene; ss.
OS	Homo sapiens.
TH	Key Location/Qualifiers
FT	cds 1..2445
FT	/tag= a
FT	/product= Peptide which modulates GDP exchange in
FT	the p21-GDP complex.
FT	cds 445..2445
FT	/tag= b
FT	/product= Peptide which modulates GDP exchange in
FT	the p21-GDP complex.
FT	cds 976..2445
FT	/tag= c
FT	/product= Peptide which modulates GDP exchange in
FT	the p21-GDP complex.
PB	MO9321314-A.
PN	28-OCT-1993.
PF	19-APR-1993; F00382.
PR	21-APR-1992; FR-004827.
PA	(RHON) RHONE POULENC RORER SA.
PI	Schweighoffer F, Tocque B;
RP	WP1: 93-351724/44

DR P-PSDB; R43578, R50990, R50991.
 PT New peptide(s) modulating GDP exchange in complexes - with ras
 PT protein and derived antibodies, nucleic acid etc. esp. for
 PT diagnosis and treatment of cancer.
 PS Claim 5; Page 16-20; 46pp; French.
 CC The peptide(s) encoded by the sequence antagonise the interaction of
 CC GDP-exchange factor (GRF) with the p21-GDP complex and thereby
 CC regulate the activity of ras gene products. They can also be used
 CC to identify other compounds which can modulate GDP exchange. The
 CC coding sequence can be used to produce antisense products which can
 CC inhibit oncogene expression and for use in diagnosis e.g. detecting
 CC (over)expression of GRF for typing of cancers.
 SQ Sequence 2652 BP; 706 A; 757 C; 629 G; 560 T;

Query Match 1.2%; Score 35.4; DB 1; Length 2652;
 Best Local Similarity 42.5%; Pred. No. 2.8;
 Matches 247; Conservative 0; Mismatches 331; Indels 3; Gaps 1;

QY 2220 ctttgaactatgatctccaaagattagatcccgatgacattatgatgggaac 2279
 DB 1721 CTTTGAACCACTCACTGAGATCGCGAGACGCTGACCTCTAGATCACCTCG 1780
 QY 2280 tcttcaactgctgatgagctgagctaatctatcacacattggaaggcataattta 2339
 DB 1781 TCTTCAGAAATCTCTTATGAGAGATTCTTCGACACAGATGATGAACTGGAAGA 1840
 QY 2340 aaaaaccacacagaacttgattgtctcgtgagagatttaaatcaattcagtttggg 2399
 DB 1841 ATGAAGAGACCCCTTATATCATGAAACCACTAAGCACTTCATGACATCAGTAACCTGA 1900
 QY 2400 tctgcaactgagatctgcttctgtctcagctcagaacgctgctcattataaaaaat 2459
 DB 1901 TTGCTTCAAGAAATATTCGCAATGAGACATCAACGCCAGGTCAGCGCATCGAGAAGT 1960
 QY 2460 ttatttaagatcagcagccacgttaagagataaaatctgaattcttcttccatcg 2519
 DB 1961 GGGTGGCCGTAGCTACATATGCGCCCTCCACAACTACAAATGCCGTATGAGATCA 2020
 QY 2520 tcatgagactaagtaacatgctgtagcgcttggcaactgtaggagaactgcca 2579
 DB 2021 CCTGTCATGAAACCGCATCTCTCCGGCTCAAAAGACGTGGCTCAAAATCTCTTA 2080
 QY 2580 gcaagttcaagaagttcatgctgagatttgaagttaatgaccttcaagaacaca 2639
 DB 2081 AGCAGACTAAAGCTTTGATTATTAAGCTCCAAAGCTTGTCTCATCTGAGGGCAGATTTA 2140
 QY 2640 gggcctacaagctgacgtagtaagctggaacctctctcatcccttcatgcttgc 2699
 DB 2141 AGAATCTCAGAGAACCTTTGAAAATTTGTGAACCCCTGTGTCTTACTGGGAGATGT 2200
 QY 2700 tcaataagatatgacattactctcatgaggggacaagaagcttca---ttgacaactag 2756
 DB 2201 ACCCTACCGACCTGCGCTTCATCGAGAGGGAGCCCAATTTACAGGAAACGCGCTGG 2260
 QY 2757 taaacttggaaaaaatgctgctgatgtaactaagcgccaga 2797
 DB 2261 TCAACTCTTCNAAGATGAGGATGATATCCATTATCCGA 2301

RESULT 6
 N80860
 ID N80860 standard; cDNA; 9601 BP.
 AC N80860;
 DT 15-NOV-1990 (first entry)
 DE Sequence corresp. to the genome of SIV isolate STLV-III Mac 142-83
 DE in clone SIV-1
 KW PSIV-1.1; PSIV-1.2; STLV-III Mac 142-83; ss.
 OS Simian immunodeficiency virus.
 FH key Location/Qualifiers
 FT cds 551..2068
 FT /tag= a
 FT /gene="gagd"

FT cds 1726..4893
 FT /tag= b
 FT /gene="pol"
 FT 4826..5467
 FT cds
 FT /tag= c
 FT /gene="g"
 FT 5298..5633
 FT /tag= d
 FT /gene="x"
 FT 5637..5939
 FT /tag= e
 FT /gene="r"
 FT 5788..6084
 FT /tag= f
 FT /gene="tail"
 FT 6014..6130
 FT cds
 FT /tag= g
 FT /gene="art1"
 FT 6090..8297
 FT /tag= h
 FT /gene="env"
 FT 8298..8732
 FT /tag= i
 FT /gene="env ctd."
 FT 8296..8391
 FT cds
 FT /tag= j
 FT /gene="tat2"
 FT 8294..8548
 FT cds
 FT /tag= k
 FT /gene="art2"
 FT 8569..9354
 FT cds
 FT /tag= l
 FT /gene="p"
 FT
 PN W0805440-A.
 PD 28-JUL-1988.
 PF 15-JAN-1988; F00025.
 PR 15-APR-1987; FR-005398,
 PA (INSP) Inst Pasteur(Aliz/).
 PI Alizon M, Montagnier L, Guetard D, Clavif F, Sonigo P, Guyader M,
 PI Tiollais P, Chakrabarti L, Desrosiers R;
 DR WPI; 88-220290/31.
 DR P-PSDB; P80802, p81783, p81784, p81785, p81691, p81756, p81757, p81758,
 DR p81759, p81760, p81761, p81762.
 PT New peptide(s) with immunological properties of HIV-2 envelope protein -
 PT having the structure of simian immune deficiency virus proteins,
 PT useful in diagnosis and of vaccine components
 PS Claim 23; Fig 1B; 86pp; French.
 CC n80860 is contained in lambda SIV-1 which was deposited on 15/4/87 at
 CC the CNRM under numbers I-658 (PSIV-1.1) and I-659 (PSIV-1.2). It is the
 CC cDNA to SIV-1 genomic RNA. Recombinant DNA contg. all or part of the cDNA
 CC from this sequence inserted into a vector are claimed. Labelled fragments
 CC of cDNA can be used as hybridisation probes. New peptides which have
 CC immunological properties in common with those of the peptide structure of
 CC the envelope protein of HIV-2 and also have a peptide structure in common
 CC with that of SIV-1 glycoprotein are claimed. Antigenic and immunogenic
 CC conjugates contg. the peptides and a kit to detect HIV-2 in biological
 CC fluids are new. The peptides are useful for in vitro diagnosis of HIV-2
 CC infection and some of them can be used as components of immunogens and
 CC vaccines against HIV. Antibodies raised against them can be used for
 CC treatment of AIDS.
 SQ Sequence 9601 BP; 3277 A; 1809 C; 2378 G; 2136 T; 1 Others;

Query Match 1.2%; Score 35.4; DB 1; Length 9601;
 Best Local Similarity 59.4%; Pred. No. 6.3;
 Matches 60; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1796 ttgagaagatttcaagcaatctcagaagaagcacaagcacaagaagcacaag 1855
 DB 8035 TTGAGGCAATATATACGGCCCTCTGAGAGAGCACAAATTAACAGAGAAACATG 8094
 QY 1856 gtctttgcaacagttcaatacggcgatgagagagccca 1896

DB 8095 TATGAAATTACAAAAGTTGATAGCTGGAGTGTGGCAA 8135

RESULT 7

ID V04063 standard; cDNA; 5183 BP.

AC V04063;

DT 08-JUN-1998 (first entry)

DE Schizosaccharomyces pombe RNA-binding protein ZPR1 cDNA.

KW ZPR1; RNA-binding protein; non-activated receptor complex;

KW signal transduction; epidermal growth factor receptor;

KW platelet derived growth factor receptor; cancer; tumour; marker;

KW angiogenesis; diabetic retinopathy; rheumatoid arthritis;

KW psoriasis; coronary atheroma; renal failure; gene therapy; ss.

OS Schizosaccharomyces pombe.

FH Key

FT CDS location/Qualifiers

FT 1076..2456

FT /*tag= a

PN W09746684-A1.

PD 11-DEC-1997.

PF 06-JUN-1997; U09911.

PR 06-JUN-1996; US-019219.

PA (UYMA-) UNIV MASSACHUSETTS.

PI Davis RJ, Galcheva-Gargova Z;

DR WPI; 98-042192/04.

DR P-PSDB: W38456.

PT Nucleic acid sequence encoding RNA-binding polypeptide ZPR1 - useful

PT to reduce angiogenesis, increase cell proliferation or kidney cell

PT regeneration or inhibit tumour growth

PS Claim 9; Fig 6; 88pp; English.

CC This nucleotide sequence comprises a cDNA clone coding for

CC Schizosaccharomyces pombe ZPR1 (see W38456), a novel protein that

CC belongs to a novel class of signalling molecules (see also

CC W38454-57) that bind to non-activated receptors (e.g. EGF and PDGR

CC receptors) and specifically bind small nucleolar RNAs (e.g. U3).

CC Saccharomyces cerevisiae, S. pombe and human ZPR1 clones (see

CC V04062-64) were identified using a mouse ZPR1 clone (see V04061) as

CC probe. The isolated clones can be expressed in the usual host

CC cells, and is a marker for growth and differentiation processes,

CC including malignant transformation of cells. ZPR1, optionally

CC expressed by gene therapy methods, can be used to inhibit the

CC proliferation of ZPR1 associated malignant cells, or for the

CC detection of ZPR1 suppressors and modulators of ZPR1 gene

CC expression. Suppressors increase cell proliferation, reduce

CC angiogenesis, e.g. to treat tumours, diabetic retinopathy,

CC rheumatoid arthritis, psoriasis and coronary atheroma, and/or

CC increase tubular regeneration of kidney cells, e.g. following acute

CC renal failure.

SQ Sequence 5183 BP; 1629 A; 906 C; 993 G; 1655 T;

Query Match 1.2%; Score 35.2; DB 1; Length 5183;

Best Local Similarity 54.7%; Pred. No. 4.9; Indels 0; Gaps 0;

Matches 70; Conservative 0; Mismatches 58;

QY 2578 aagcaagttcaagaagttatcgagattgaagtttaagacccttcaagagaca 2637

DB 1970 AAGACTGGTGTGAATTCACCCAAAGTCGAAATAATTAAGTCAATGGATGCC 2029

QY 2638 caggcgctacagctgacagtagctaagctggaacctctcatcccttaatgctt 2697

DB 2030 GAGGACTTATCCCGTATATTCGAATCTGAACCGCATCTCTTAAATVCTGTAACT 2089

QY 2698 gctcatta 2705

DB 2090 GGACTTGA 2097

RESULT 8

ID T37405 standard; cDNA; 2128 BP.

AC T37405;

DT 26-NOV-1996 (first entry)

DE Prostate tumour inducing gene PTI-1.

KW Prostate tumour inducing gene-1; PTI-1; oncogene;

KW elongation factor-1 alpha; diagnosis; therapy; metastasis;

KW cell surface expression; antigen; cancer; ss.

OS Homo sapiens.

FH Key

FT primer_bind location/Qualifiers

FT complement (317..336)

FT /*tag= a

FT /*note= "L primer"

FT primer_bind 567..586

FT /*tag= b

FT /*note= "A primer"

FT 621..1817

FT cds

FT /*tag= c

FT /*transl_except= (744..746, aa:Gly)

FT /*note= "ACC encodes Thr"

PN W09621671-A1.

PD 18-JUL-1996.

PF 11-JAN-1996; U00307.

PR 11-JAN-1995; US-371377.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Fisher PB, Shen R;

DR WPI; 96-342235/34.

DR P-PSDB: W03518.

PT DNA encoding prostate carcinoma tumour antigen and prostate tumour

PT inducing genes - and related proteins, vectors, antibodies, etc.,

PT for diagnosis and treatment of metastatic cancer

PS Disclosure; Fig 8A; 169pp; English.

CC Prostate tumour inducing gene-1 (PTI-1) (T37405) is a novel putative

CC oncogene that may contribute to carcinoma development in human

CC prostate and other tissues. PTI-1 was initially identified in

CC human prostatic carcinoma LNCaP cell DNA-transfected tumour-derived

CC murine CRE-Trans 6 cells using an RNA differential display method,

CC and full-length cDNA can be directly cloned from an LNCaP cDNA

CC library. The PTI-1 gene comprises a unique 5' 630 bp region and

CC a 3' truncated and mutated elongation factor-1 alpha gene. It

CC codes for a 46 kDa protein (W03518). PTI-2 (T37412), PTI-3 (T37413)

CC and a prostate carcinoma tumour antigen gene (T37414) were also

CC identified. The PTI-1 gene can be used to design probes useful

CC e.g. for detection of metastatic cancer, or to produce recombinant

CC proteins. Antisense sequences can be used in cancer therapy.

SQ Sequence 2128 BP; 608 A; 434 C; 536 G; 550 T;

Query Match 1.1%; Score 34.6; DB 1; Length 2128;

Best Local Similarity 54.3%; Pred. No. 4.3;

Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 100 gttattatgagaatctggaagaaggaataacattatctgcacaggtatattgaaaca 159

DB 1573 GTTCTGGTAAAGCTGGAAGATGGCCCTTAATTTCTGAAGTGTGATGCTGCATTG 1632

QY 160 ctgtagatgctgctgcagaggtcttgatgattgaagatctgagacagagcagcag 219

DB 1633 TTGATATGCTTCTCGCAAGCCCATGCTGTGTGAGAGCTTCTCAGACTATTCACCTTTGG 1692

QY 220 gatgctgctg 228

DB 1693 GCTGCTTGG 1701

RESULT 9

ID Q14183 standard; cDNA; 2338 BP.

AC Q14183;

DT 15-JAN-1992 (first entry)

DE N.clavipes dragline silk protein coding sequence.

KW protein super fibre; major ampullate silk; orb web spider; ss.

OS Nephilila clavipes.

FH Key

FT cds location/Qualifiers

FT 1..2157

FT /*tag= a

```
FT      EP-452925-A.                               /product= silk protein 1
PN      PD 23-OCT-1991.
PE      PE 18-APR-1991. 106217.
PR      PR 20-APR-1990: US-511792.
PA      (UYWY-) UNIV OF WYOMING.
PI      Lewis RV, Xu M, Himman M;
DR      WPI: 91-312199/43.
P-PSDB: R14308.
PT      DNA encoding spider silk protein-1 and 2 and variants - isolated
PT      from Nephila clavipes, for prodn. of spider silk protein and
PT      fibres having desired characteristics
PS      Claim 4; Page 23; 48pp; English.
CC      A.N.clavipes major ampullate gland cDNA library was screened with
CC      probes based on peptide fragments of the purified spider silk
CC      protein (see Q14185). Positive plaques were identified and the
CC      spider silk protein 1 coding sequence was determined. See also
CC      Q14184.
SQ      Sequence 2338 BP; 566 A; 433 C; 916 G; 423 T;

Query Match      1.1%; Score 34.6; DB 1; Length 2338;
Best Local Similarity 50.3%; Pred. No. 4.5;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY      786 ttctgatgatgagcagcagatgcccccttgcctactgagagagaagaagatgctg 845
DB      435 TGGTGTGACAAAGAGATATGAGGCTTGGAAAGCCAAAGCAGACGAGATGATTAGG 494
QY      846 atgagagatcctcagacacccatctgtctgtctgtacagatgagcccgagccacatga 905
DB      495 TGGACAAAGCTGACAGTGCAGCAGCAGCAGCCGAGGCTCTGACAAAGCGGATACGG 554
QY      906 gtagatcctctgcgaaccactgctgtctgtctgtacagatgagcccgagccacatga 954
DB      555 TGGCTTGTGTGACAAAGTGCAGCAGCAGCAGCGGAGGCTCTGACAAAGCGGATACGG 603

RESULT 10
V23249
ID      V23249 standard; cDNA; 2338 BP.
AC      V23249;
DE      06-JUN-1998 (first entry)
KW      Nephila clavipes spider silk protein encoding cDNA.
KW      Spider: Nephila clavipes; silk protein; tandem repeat; fibre; dragline;
KW      cocoon; tensile strength; elasticity; ss.
OS      Nephila clavipes.
FH      Key      Location/Qualifiers
FT      CDS      1..2157
FT      FT      /*tag= a
FT      FT      /product= "spider silk protein 1"

PN      US5728810-A.
PD      17-MAR-1998.
PE      PE 19-APR-1995: 425069.
PR      PR 15-APR-1991: US-684819.
PR      20-APR-1990: US-511792.
PR      04-OCT-1994: US-517844.
PR      19-APR-1995: US-425069.
PA      (UYWY-) UNIV WYOMING.
PI      Himman MB, Lewis RV, Xu M;
DR      WPI: 98-270437/24.
P-PSDB: W53346.
PT      Recombinant spider silk proteins - useful for making fibres
PT      Example 3; Column 25-30; 68pp; English.
CC      The present sequence encodes a spider silk protein from the present
CC      invention. Spider silk proteins, and peptide fragments within the
CC      proteins, can be produced and purified independently and can then be
CC      mixed and made into fibres that have higher tensile strengths and
CC      elasticity than naturally occurring fibres. The fibres can be used in
CC      mixed composites. The invention allows the two naturally occurring
CC      Nephila clavipes silk proteins to be produced independently so that
CC      they can later be combined to form silk fibres of high tensile strength
CC      and elasticity.
```

```
SQ      Sequence 2338 BP; 564 A; 432 C; 918 G; 424 T;

Query Match      1.1%; Score 34.6; DB 1; Length 2338;
Best Local Similarity 50.3%; Pred. No. 4.5;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY      786 ttctgatgatgagcagcagatgcccccttgcctactgagagagaagaagatgctg 845
DB      435 TGGTGTGACAAAGAGATATGAGGCTTGGAAAGCCAAAGTGCAGCAGAGGTGATTAGG 494
QY      846 atgagagatcctcagacacccatctgtctgtctgtacagatgagcccgagccacatga 905
DB      495 TGGACAAAGTGCAGTGCAGCAGCAGCAGCCGAGGCTCTGACAAAGCGGATACGG 554
QY      906 gtagatcctctgcgaaccactgctgtctgtctgtacagatgagcccgagccacatga 954
DB      555 TGGCTTGTGTGACAAAGTGCAGCAGCAGCAGCGGAGGCTCTGACAAAGCGGATACGG 603

RESULT 11
V65139/c
ID      V65139 standard; DNA; 4833 BP.
AC      V65139;
DE      02-MAR-1999 (first entry)
DE      Ehrlichia sp. W20 genomic DNA.
KW      Granulocytic ehrlichia; GE; W20; tick-borne infection; fatal; vaccine;
KW      Immune response; detection; diagnosis; Ehrlichiosis; ss.
OS      Ehrlichia sp.
FH      Key      Location/Qualifiers
FT      CDS      42..1274
FT      FT      /*tag= a
FT      FT      /product= "W20.1"
FT      FT      2059..3603
FT      FT      /*tag= b
FT      FT      /product= "W20.2"

PN      M09849312-A2.
PD      05-NOV-1998.
PE      PE 24-APR-1998: U08264.
PR      PR 25-APR-1997: US-044869.
PA      (AQUILA-) AQUILA BIOPHARMACEUTICALS INC.
PI      Beltz G, Coughlin RT, Murphy C, Storey J;
DR      WPI: 99-034663/03.
P-PSDB: W81975, W81976.
PT      New isolated granulocytic ehrlichia nucleic acids - used to develop
PT      products for use in vaccines for inhibiting Ehrlichiosis and for use
PT      in detection and diagnosis
PS      Claim 2; Fig 1; 184pp; English.
CC      This sequence encodes two proteins which are associated with
CC      granulocytic ehrlichia (GE). W20.1 and W20.2 which have been isolated
CC      from H160 cells infected with Ehrlichia sp. GE is an acute potentially
CC      fatal tick borne infection and the proteins described in this invention
CC      can be used in vaccines to elicit a beneficial immune response in an
CC      animal to GE. They can be used for inhibiting Ehrlichiosis in an animal.
CC      The products can also be used for detection and diagnosis.
SQ      Sequence 4833 BP; 1484 A; 797 C; 1114 G; 1438 T;

Query Match      1.1%; Score 34.6; DB 1; Length 4833;
Best Local Similarity 56.6%; Pred. No. 7.1;
Matches 64; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY      1085 gaagaagtgatcctccctggtcatattcttaagaagatcagtgatgtatcattacggc 1144
DB      4108 GAAAAGCATCTCCCTCCAGGTTTACGTAACTTACAGAGCTAGATAAATAATTACGTCAATC 4049
QY      1145 aagggttggtctgcacccctgcataagaagatgactcgcaagttagcact 1197
DB      4048 AGGATATGCTTACAACTTCATTAAGTGAATGATGTAGTAAGAAACACT 3996

RESULT 12
V74475
```

FT V74475 standard; DNA: 22243 BP.
AC V74475;
DE 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #164.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PH Key
FT misc-feature
FT Location/Qualifiers
FT 61..120
FT /tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT 1861..1920
FT /tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 3661..3720
FT /tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 5461..5520
FT /tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 7261..7320
FT /tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 9061..9120
FT /tag= f
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 10861..10920
FT /tag= g
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 12661..12720
FT /tag= h
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 14461..14520
FT /tag= i
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 16261..16320
FT /tag= j
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 18061..18120
FT /tag= k
FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 19861..19920
FT /tag= l
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature
FT 21661..21720
FT /tag= m
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT EP-766519-A2.
FT 30-JUL-1997.
FT 07-JAN-1997; 100117.
FT 05-JAN-1996; US-009861.
FT (HUMA-) HUMAN GENOME SCI INC.
FT Barash SC, Choi GH, Dillon PJ, Pannon MR, Kunsch CA,
FT Rosen CA;
FT WPI; 97-374922/35.
FT Polynucleotide(s) and proteins derived from Staphylococcus aureus
FT stored on computer readable medium and used in the production of
FT anti-S.aureus vaccines
FT Claim 1; Page 800-813; 3271pp; English.
FT This sequence represents one of 5191 Staphylococcus aureus DNA sequences
FT of the invention. The DNA sequences are recorded on a computer readable
FT medium, preferably selected from a floppy or hard disk, random access
FT memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
FT the S.aureus DNA sequences allows putative functions to be assigned so
FT that protein-encoding or regulatory regions of commercial, therapeutic or
FT industrial importance can be obtained. Specifically, sequences which are
FT likely to encode antigens have been identified and these polypeptides can
FT be used in a vaccine composition against S.aureus infection. The
FT CC polypeptides can also be used in a kit for the immunodetection of
FT S.aureus in a sample. S.aureus is implicated in numerous human diseases,
FT including cellulitis, eyelid infections, food poisoning, osteomyelitis,
FT skin and surgical wound infections, scalded skin syndrome, toxic shock
FT syndrome, etc. Organisms transformed with the DNA sequences can be used
FT CC for recombinant production of the polypeptides. The new DNA sequences
FT CC (and their fragments) are useful as primers or probes for isolating
FT CC homologues of any of the S.aureus DNA sequences contained on the
FT CC computer readable medium.
FT Sequence 22243 BP; 7693 A; 2910 C; 4245 G; 6604 T;
SQ
Query Match 1.1%; Score 34.6; DB 1; Length 22243;
Best Local Similarity 51.6%; Fred. No. 18;
Matches 79; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1 gatccagcgaagtgtgatataatcttcactcgcagcgaagaagttaagcttgaga 60
DB 242 GCTTCATCACAACAGAGTTTAATAGTATCATAGATGAACATATATTAACAATATT 301
QY 61 aatttcaccaaatcttccttcatacgaattgcttaatgtgttatattgaaatcgtgaa 120
DB 302 AATGACAAAAGACATCTTCATCCAGAGTTGTAATTCGATTCAGAGAACAAAGAA 361
QY 121 aggaataacatatttcgccagcggtgatatg 153
DB 362 AAGTAATTTCTTGATTTGGCAAGTTGTGTATG 394
RESULT 13
V06059
ID V06059 standard; cDNA, 1677 BP.
AC V06059;
DE 01-MAY-1998 (first entry)
DE Human imidazole receptor subtype 1 cDNA.
KW Imidazole receptor subtype 1; IRL; screening; antagonist; agonist;
KW antibody; C-terminal fragment; ds.


```

Oy 258 DEGLKALVNDAPRAASIVLREDNCHFLRVYKEDFNRLRDVEANTVRLKEHDQDVLVLEK 317
Db 480 VAGNARNASOGNSOPQOKTVMSTGPEKLEHEFTIRLEATLNATOSVLDNFTMHCY 539
Oy 318 VAGNARNASOGNSOPQOKTVMSTGPEKLEHEFTIRLEATLNATOSVLDNFTMHCY 377
Db 540 FMPNTQLCPALVAHYHAPOSGTDEQEKMDYALNNKRVIRLVLMQAAVMDLQEDDDVSM 599
Oy 378 FMPNTQLCPALVAHYHAPOSGTDEQEKMDYALNNKRVIRLVLMQAAVMDLQEDDDVSM 437
Db 600 ALEEFYVSVSDARMAIALKEQLPELEKIVQISEDAKAPQKKHVLLOQFNTGDERAQ 659
Oy 438 ALEEFYVSVSDARMAIALKEQLPELEKIVQISEDAKAPQKKHVLLOQFNTGDERAQ 497
Db 660 KROPINGSDEVLFKYCMQDHTYTTIRVPATSVKEVISAVDKLGSGGLTIYKMSGGE 719
Oy 498 KROPINGSDEVLFKYCMQDHTYTTIRVPATSVKEVISAVDKLGSGGLTIYKMSGGE 557
Db 720 KYVLKPNDSVFTTLTLINGRLFACPREQFDSLTPLEQEGPTVGTGFEILMSSKDLYQ 779
Oy 558 KYVLKPNDSVFTTLTLINGRLFACPREQFDSLTPLEQEGPTVGTGFEILMSSKDLYQ 617
Db 780 MIIYMELEFNCVHELELIYHFGHNFKKTANDLFLRRENIQFWVTEICLSQSLK 839
Oy 618 MIIYMELEFNCVHELELIYHFGHNFKKTANDLFLRRENIQFWVTEICLSQSLK 677
Db 840 RYQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIAVSRLATLWELPSKFKFAEESL 899
Oy 678 RYQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIAVSRLATLWELPSKFKFAEESL 737
Db 900 MDSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKNFTIDNLVNEFKRMJANTAR 959
Oy 738 MDSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKNFTIDNLVNEFKRMJANTAR 797
Db 960 TVRYRSPQPNDAQAANKNHODVSYROLVINDNQRTLSOMSHLEPRRP 1011
Oy 798 TVRYRSPQPNDAQAANKNHODVSYROLVINDNQRTLSOMSHLEPRRP 849

RESULT 2
ID 0921P0 PRELIMINARY; PRT; 993 AA.
AC 0921P0;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
RT "A brain CAMP-dependent Rap1 guanine-nucleotide exchange factor.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DE CAMP-DEPENDENT RAP1 GUANINE-NUCLEOTIDE EXCHANGE FACTOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA GAUDRIAULT G.E., TAKAYA K., VALE W.W.;
RT "A brain CAMP-dependent Rap1 guanine-nucleotide exchange factor.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR HSSP; P00515; 2BPk.
SQ SEQUENCE 993 AA; 113488 MW; 52059F9A CRC32;

Query Match 93.8%; Score 5835; DB 11; Length 993;
Best Local Similarity 93.5%; Pred. No. 0.00e+00;
Matches 795; Conservative 28; Mismatches 26; Indels 1; Gaps 1;

```

```

Oy 120 DFYLFYRLDEHEDAPLPTFEKKCEDEELQDTMLLSQMGDPAHMRLKRPQGRV 179
Db 324 DDELIYDELIHIALSHLSTTVAKRELAVLIFESNAKGSTVLFNOEGESTWYILKGS 393
Oy 180 DDELIYDELIHIALSHLSTTVAKRELAVLIFESNAKGSTVLFNOEGESTWYILKGS 239
Db 384 VNVYTKGVYCTLHEDDDDGKALVNDAPRAASIVLREDNCHFLRVYKEDFNRLRDVE 443
Oy 240 VNVYTKGVYCTLHEDDDDGKALVNDAPRAASIVLREDNCHFLRVYKEDFNRLRDVE 299
Db 444 ANTVRLKEHDQDVLVLEKVPAGNARNASOGNSOPQOKTVMSTGPEKLEHEFTIRLEPS 503
Oy 300 ANTVRLKEHDQDVLVLEKVPAGNARNASOGNSOPQOKTVMSTGPEKLEHEFTIRLEPS 359
Db 504 LNEATDVLNDFVMHCVFMPNTQLCPALVAHYHAPOSGTDEQEKMDYALNNKRVIRLV 563
Oy 360 LNEATDVLNDFVMHCVFMPNTQLCPALVAHYHAPOSGTDEQEKMDYALNNKRVIRLV 419
Db 564 LQNAAMTGDLLQEDDDVMAFLPEFYVSVSDARMAIALEQLPELEKIVQISEDAKAPQ 623
Oy 420 LQNAAMTGDLLQEDDDVMAFLPEFYVSVSDARMAIALEQLPELEKIVQISEDAKAPQ 479
Db 624 KKHVLLQDQNTGDERAQKROPTRGSDEVLFKYCIDHTYTTIRVPAASVKEVISAVD 683
Oy 480 KKHVLLQDQNTGDERAQKROPTRGSDEVLFKYCIDHTYTTIRVPAASVKEVISAVD 539
Db 684 KLSGSEGLIYKMSGGEKVVYKSNDSVFTTLTLINGRLFACPREQFDSLTPLEQEGPT 743
Oy 540 KLSGSEGLIYKMSGGEKVVYKSNDSVFTTLTLINGRLFACPREQFDSLTPLEQEGPT 599
Db 744 TGTVGFPELMSKDLYQMTTYDMELFNCVHELELIYHFGHNFKKTANDLFLRPN 803
Oy 600 TGTVGFPELMSKDLYQMTTYDMELFNCVHELELIYHFGHNFKKTANDLFLRPN 659
Db 804 EIOFWVTEVCLQSOLSKRYQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIAVSRLAT 863
Oy 660 EIOFWVTEVCLQSOLSKRYQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIAVSRLAT 719
Db 864 WEKLPSEKFKFYAEFSLMDPSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKNFT 923
Oy 720 WEKLPSEKFKFYAEFSLMDPSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKNFT 779
Db 924 IDNLVNEFKRMJANTARVRYRSPQPNDAQAANKNHODVSYROLVINDNQRTLSQ 983
Oy 780 IDNLVNEFKRMJANTARVRYRSPQPNDAQAANKNHODVSYROLVINDNQRTLSQ 839
Db 984 MSHRLEPRRP 993
Oy 840 MSHRLEPRRP 849

RESULT 3
ID 0921C7 PRELIMINARY; PRT; 436 AA.
AC 0921C7;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
RT "A family of CAMP-binding proteins that directly activate Rap1.";
RL EMBL; U78517; AAD03423.1; -.
DR EMBL; U78517; AAD03423.1; -.
FT NON_TER 1
SQ SEQUENCE 436 AA; 50117 MW; 30F35FB2 CRC32;

```

Query Match 49.1%; Score 3052; DB 11; Length 436;
Best Local Similarity 95.6%; Pred. No. 0.00e+00;
Matches 417; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

```
Db 1 RVIRLVYQMAAMGDLQEDVDVAMALEEYVSSDDARMMAVAFKQALAEKTVQKQISE 60
QY 414 RVIRLVYQMAAMGDLQEDVDVAMALEEYVSSDDARMMAVAFKQALAEKTVQKQISE 473
Db 61 DAKAPQKKHVVLLQOQFNGDERAKROPINGSDVLEFKVYCIDHTPTTIRVPAASVKEV 120
QY 474 DAKAPQKKHVVLLQOQFNGDERAKROPINGSDVLEFKVYCIDHTPTTIRVPAASVKEV 533
Db 121 ISAVADKLGSSEGLIIVKMSNGEKVVLKPNDSVFTTLINGRLPACPREQDSITPLP 180
QY 534 ISAVADKLGSSEGLIIVKMSNGEKVVLKPNDSVFTTLINGRLPACPREQDSITPLP 593
Db 181 EQEGPTTGYGTGTEFLMSKDLAVQMTTYDMLFNCVLELELIYHFGRNHFKKTTANL 240
QY 594 EQEGPTTGYGTGTEFLMSKDLAVQMTTYDMLFNCVLELELIYHFGRNHFKKTTANL 653
Db 241 FLRFENIOWFVYETICLSOLSKRVOLKKCIAHCKEYKNLSFPGIVAGLSNVAE 300
QY 654 FLRFENIOWFVYETICLSOLSKRVOLKKCIAHCKEYKNLSFPAIVAGLSNIAV 713
Db 301 SRLATWEKLPKSKFKFYAFESIMDPSRNHKAAYRLTAKEPLLPMPMLLKDMTFPH 360
QY 714 SRLATWEKLPKSKFKFYAFESIMDPSRNHKAAYRLTAKEPLLPMPMLLKDMTFPH 773
Db 361 EGKNTFTDNLVNEKEMIMANTARTVYRYSQFPNDPAQANKNHDVRSYVQLNVIND 420
QY 774 EGKNTFTDNLVNEKEMIMANTARTVYRYSQFPNDPAQANKNHDVRSYVQLNVIND 833
Db 421 ORTISOMSHLEPRRP 436
QY 834 ORTISOMSHLEPRRP 849
```

```
RESULT 4
ID 095398; PRELIMINARY; PRT; 881 AA.
AC 095398;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE RAP1 GUANINE-NUCLEOTIDE EXCHANGE FACTOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RX MEDLINE; 99068645.
RA DE ROOIJ J., ZWARTKRUIS F.J.T., VERHEIJEN M.H.G., COOL R.H.,
RA NUDMAN S.M.B., WITTINGHOFF A., BOS J.L.;
RT "Epac is a Rap1 guanine-nucleotide-exchange factor directly activated
RT by cyclic AMP.";
RL Nature 396:474-477(1998).
DR EMBL; A0103905; AAC83381.1; -.
DR HSP; P00515; 2BPk.
SQ SEQUENCE 881 AA; 99312 MW; 43A29191 CRC32;
```

Query Match 47.3%; Score 2942; DB 4; Length 881;
Best Local Similarity 51.1%; Pred. No. 0.00e+00;
Matches 406; Conservative 189; Mismatches 188; Indels 11; Gaps 8;

```
Db 30 RWPPLINSESLFSESLEQASTERYVLRAQQLRHRLATCPNLTRDKRYHLRYQCS 89
QY 16 RRTHLIPHPVLRPANTITVPSKIRLAKGILRNALISAPRMIDRKXHLTYQCCV 75
Db 90 GRELVYDILALGLGYSRSOVGICQVLDLDEGALCHYKHMAQODRACQYRRPGEPE 148
QY 76 GTELVDMVIDETPCVHSRTQAVGMQVLLDGVNLHVDQEHHPQDF-YLTYRFLDDEHD 134
```

```
Db 149 -PVGTHE--ME--ETIAEAVALLSQRGPDALTVLRKPGQRTDEDLIFEELHIKA 203
QY 135 APLPFEKKEDEDELQDTMLLSQMGPDAMHMLLRKPGQRTDDELIITEELHIKA 194
Db 204 VAHLSNVAKRELAAILPEPSSKAGTVLFSGDKSTWYIIWKGSVNVYHKGGLVTLH 263
QY 195 LSHLSTVARELAGVLIFFESHAKGTVLFNQEESTSWYIIKGSVNVYHKGGLVTLH 254
Db 264 EGDDEGQALVNDAPRAATIIREDNCHFLRYDKODPFRRIIKDVAKMRLREKRVLV 323
QY 255 EGDDEGQALVNDAPRAATIIREDNCHFLRYDKODPFRRIIKDVAKMRLREKRVLV 314
Db 324 LERASGA-GPSPPTPGNRRTVMSGTEPEKLELLEPAMGPDSSAHPTPEFLSDFLT 382
QY 315 LKVPAGNNAASNOGSOPOQKTVMSGTEPEKLELLELIRLEALNATSDVINDFTM 374
Db 383 HVEFMSAOLCAALLHHFHPVAGSGEQRSTYVCNKRQOILRLVSQWVALYGSMLHTDP 442
QY 375 HCVFMENTQLCPALVAHYHAQPSQGTQEQEKMDYALNNKRRVIRLVQMAAMGDLQEDD 434
Db 443 VATSFLQKISDLVGRDTRLNLRQWPERRCHGLECCGASQMAKRNLPWLPNOD 502
QY 435 VMAFLFEYVSSDDARMIAALKEDLPLEKTVQKQISE-DAKAPQKKHVVLLQOQFNG 493
Db 503 EPLPGSSCAIOGVGDVYPDICTPDSVLTLOPVASREVMAALAOEDGNTKGOVLKY 562
QY 494 ERAQKRQ-PIRSDEVLEFKVYCIDHTPTTIRVPAASVKEVISAADKLGSSEGLIIVKM 552
Db 563 NSAGDAIGLQDPARKVATSLGNERLFFVNPQEAHELIPHPDOLGPTVSAEGDLIVSAK 622
QY 553 SSGGEKVVLPKPNDSVFTTLINGRLPACPREQDSITPLPQEGPTTGYGTGTEFLMSK 612
Db 623 DIAGOLTHDMSLFNSIHQVELIHYVLPQHLRDVTTANLEFMRNRELDQWATVETCL 682
QY 613 DLAYQMTTYDMLFNCVHLELIYHFGRNHFKK-TTANLDELFRFENIOWFVYETICL 671
Db 683 CPVPGPRAQLKFKTKLAHLKEOKNLSFPAVMGLNSAISRAHMERLPHVRLY 742
QY 672 CSOLSKRVOLKKFKTKIAHCKEYKNLSFPAIVAGLSNIAVSRALATWEKLPKSKFKFY 731
Db 743 SALERLDPSWNRHRYRLALAKLSPVPIPFMPLLKDMTFHESNHTLVENLINEKRM 802
QY 732 AFEESIMDPSRNHKAAYRLTAKEPLLPMPMLLKDMTFHEGKFTDNLVNEKRM 791
Db 803 MARARMLHCHSH 816
QY 792 IANTARTVYRYSQ 805
```

```
RESULT 5
ID 095634; PRELIMINARY; PRT; 881 AA.
AC 095634;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CAMP-REGULATED GUANINE NUCLEOTIDE EXCHANGE FACTOR I.
GN CAMP-GEFI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 99074384.
RA KAWASAKI H., SPRINGETT G.M., MOCHIZUKI N., TOKI S., NAKAYA M.,
RA MATSUDA M., HOUSMAN D.E., GRAYBIEL A.M.;
RT "A family of CAMP-binding proteins that directly activate Rap1.";
RL Science 282:2275-2279(1998).
DR EMBL; U78168; A012740.1; -.
DR HSP; P00515; 2BPk.
SQ SEQUENCE 881 AA; 99376 MW; 8C1C3C82 CRC32;
```

Query Match 47.2%; Score 2936; DB 4; Length 881;
Best Local Similarity 50.9%; Pred. No. 0.00e+00;

Matches 404; Conservative 190; Mismatches 189; Indels 11; Gaps 8;			
Db	30	RMPRLNSESLEDFSSLDQASTERYLRAGROLDHOLATCNLRKRYHLRKYRCCS	89
Qy	16	RRHLLRPVPLRPANTIKRVSEKILRAGKILRNALISRAPIHRIKRYHLKTYRCCV	75
Db	90	GRELVDGILALGIVNHSQAVGICQVLLDEGALCHVKHDMAFORDOAFYFPPPEPE	148
Qy	76	GTELVDMWIDETPCVHSRQAVGMQVLLDEGVLNHVDEHNHODF-YLFYFLEDEHED	134
Db	149	-PVGTH--ME--EELAEVALISQRPDALLTVALKRPPGQRTDELDLFEELHIKA	203
Qy	135	APLPTPEEKKECDELDOTMLLSOMGRDANHMILRKPPGQRTVDLEIIEELHIKA	194
Db	204	VAHLSNYSKRELAVALLEFPHSKAGTVLEFSQDKGTSWIIIMKGSVNVYTHKGLVTLH	263
Qy	195	LSHLSTYKRLAGVLLFESHAGKTVLFRNÖGEGTSWIIILKGSVNVYTHKGLVTLH	254
Db	264	EGDDFGQALVNDAPRAATITLREYNCHFLRVDKODFNRIKIDVEAKTYRLEHKGAVLV	323
Qy	255	EGDDFGKIALVNDAPRAASIVLRDNCHFLRVDKEDFNILRDVEANTVRLKEHDQVLY	314
Db	324	LEPASOGA-GPSRPTPGRRNRYTWSGTPDKILLELEAMGLDSSANDKETFLSDFLIT	382
Qy	315	LEKVPAGNANASNGNSQPOQKYTMVGTPREKILEHLEFLIRLEAVTLEATDVLNDETMM	374
Db	383	HREMPASOALCALHHPHVERPAGSEORSTYVCNKRQOILRLYSQWALGSMILHDP	442
Qy	375	HCFHMPNTQICRALVANYHNAOPSQTEQEKMDALNNKRRVRLVQWAMAMGDLLOEDD	434
Db	443	VATSFHLKSLDVGDRTRLSNLRQWPERRCHRLENGCGNASPQMKARNLPMVLPND	502
Qy	435	VSMATLEEFYVSSDARMIALKEQLRELEKIVKOISE--DAKAPQKKHVVLLQDFNGD	493
Db	503	EPLEPGSCALIQVGDKVPYDICRPHSVTLQDPYASVREYMAALAOEDGKTGQVLYK	562
Qy	494	ERAKQRO-PINGSEVLEFYVCMDHTYTTIRVPATSVKEVISAADKLGSGEGILLVM	552
Db	563	NSAGDAIGLOPDAVGATSLGNERLFVUNPOEVELIRPHROLOPBYVSGAGLDLVSK	622
Qy	553	SSGGEVVLKPNDSVFTTLTINGRLFACRPOFDSLPRDEQGYTGTGTFELMSK	612
Db	623	DLAQGLTDHWSLFSNHOVELIHVGLRPHQNRDVTANLEFRMRFNELQYVATELCL	682
Qy	613	DLAQMTIYDWELEFNCVHELELIYHFGHNFKK--TTANLDLFRRENIQGVVYTELCL	671
Db	683	CPVPGPRAQLLKFTKLAHLKEQKVNVSFPAVMFGLSNPSISLAHTWERLPHKVRKLY	742
Qy	672	CSQLSKRVOLLKFTKIIAHCKEYKNLNSFPAIVMGLSINAVSRALATWEKLPSEKFEKY	731
Db	743	SALERLDPSMNHRYRLALATKSPVIRPMLLLKDMFTIEGNTLVENLINFEKMM	802
Qy	732	AEEESLMDPSRNHRYRLVAKLEPRLPMLIKDMFTIEGNTFIDNLVNEFKMM	791
Db	803	MARARMLHCRSH 816	
Qy	792	IANTARTVRYRSQ 805	
RESULT	6	PRELIMINARY; PRT; 884 AA.	
ID	Q92IC8		
AC	Q92IC8;		
DT	01-MAY-1999 (Tremblrel. 10, Created)		
DT	01-MAY-1999 (Tremblrel. 10, last sequence update)		
DT	01-NOV-1999 (Tremblrel. 12, last annotation update)		
DE	CAMP-REGULATED GUANINE NUCLEOTIDE EXCHANGE FACTOR 1.		
GN	CAMP-GEFI.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 99074384.		

RA	KAWASAKI H., SPRINGETT G.M., MOCHIZUKI N., TOKI S., NAKAYA M.,		
RA	MATSUDA M., HOSUMI D.E., GRAVIEL A.M.;		
RT	"A family of CAMP-binding proteins that directly activate Rap1."		
RL	Science 282:2275-2279(1998).		
DR	EMBL; U78167; RAD12739.1; -.		
DR	HSSP; P00515; 2BP.		
SO	SEQUENCE 884 AA; 100256 MW; C8BB3814 CRC32;		
Query Match 45.3%; Score 2820; DB 11; Length 884;			
Best Local Similarity 50.1%; Pred. No. 0.00e+00;			
Matches 401; Conservative 186; Mismatches 200; Indels 14; Gaps 10;			
Db	30	RMPRLNSESLEDFSSLDQASTERYLRAGROLDHOLATCNLRKRYHLRKYRCCS	89
Qy	16	RRHLLRPVPLRPANTIKRVSEKILRAGKILRNALISRAPIHRIKRYHLKTYRCCV	75
Db	90	GRELVDGILALGIVNHSQAVGICQVLLDEGALCHVKHDMHTFODRDOAFYR-PE- 146	
Qy	76	GTELVDMWIDETPCVHSRQAVGMQVLLDEGVLNHVDEHNHODF-YLFYFLEDEHED	134
Db	147	-PQAG--THVEELEYEAMALLSQRPDALLTVALKSGQRTDELDLFEELHIKA	203
Qy	135	APLPTPEEKKECDELDOTMLLSOMGRDANHMILRKPPGQRTVDLEIIEELHIKA	194
Db	204	VAHLSNYSKRELAVALLEFPHSKAGTVLEFSQDKGTSWIIIMKGSVNVYTHKGLVTLH	263
Qy	195	LSHLSTYKRLAGVLLFESHAGKTVLFRNÖGEGTSWIIILKGSVNVYTHKGLVTLH	254
Db	264	EGDDFGQALVNDAPRAATITLRENNCHFLRVDKODFNRIKIDVEAKTYRLEHKGAVLV	323
Qy	255	EGDDFGKIALVNDAPRAASIVLRDNCHFLRVDKEDFNILRDVEANTVRLKEHDQVLY	314
Db	324	LEPASOGA-GPSRPTPGRRNRYTWSGTPREKILEHLEFLIRLEAVTLEATDVLNDETMM	382
Qy	315	LEKVPAGNANASNGNSQPOQKYTMVGTPREKILEHLEFLIRLEAVTLEATDVLNDETMM	374
Db	383	HREMPASOALCALHHPHVERPAGSEORSTYVCNKRQOILRLYSQWALGSMILHDP	442
Qy	375	HCFHMPNTQICRALVANYHNAOPSQTEQEKMDALNNKRRVRLVQWAMAMGDLLOEDD	434
Db	443	SDPVATSFQKSLDVSROTLSNLRQYRERRRHLRLENGCGNVSPOTKARNAPWEP	502
Qy	432	EDVSMATLEEFYVSSDARMIALKEQLRELEKIVKOISE--DAKAPQKKHVVLLQDFN	490
Db	503	NHEERLPSSAALIVGDKVPYDICRPHSVTLQDPYASVREYMAALAHEDMTGQVLY	562
Qy	491	TGDERAQRO-PINGSEVLEFYVCMDHTYTTIRVPATSVKEVISAADKLGSGEGILL	549
Db	563	VKVSAGDVVGLQPDARGVATSLGNERIFVUNPOEVELIRPHROLOPBYVSGAGLDLV	622
Qy	550	VKMSGGKVVLLKPNDSVFTTLTINGRLFACRPOFDSLPRDEQGYTGTGTFELMSK	609
Db	623	SAKDLAQGLTEHWNLEFNRIHOVELIHVGLRPHQNRDVTANLEFRMRFNELQYVATE	682
Qy	610	SSKDLAQMTIYDWELEFNCVHELELIYHFGHNFKK--TTANLDLFRRENIQGVVYTE	668
Db	683	ICLCPVGPRAQLLKFTKLAHLKEQKVNVSFPAVMFGLSNPSISLAHTWERLPHKVR	742
Qy	669	ICLCSQSKRVOLLKFTKIIAHCKEYKNLNSFPAIVMGLSINAVSRALATWEKLPSEK	731
Db	743	KLYALERLDPSMNHRYRLATKLSPPVIRPMLLLKDMFTIEGNTLVENLINFEKMM	802
Qy	729	KFYAFESLMDPSRNHRYRLVAKLEPRLPMLIKDMFTIEGNTFIDNLVNEFKMM	788
Db	803	MRMARARMLHCRSHSTAD 823	
Qy	789	MRMIANTARTVRYRSQPNP 809	
RESULT	7	PRELIMINARY; PRT; 580 AA.	
ID	Q92565		
AC	Q92565;		
DT	01-FEB-1997 (Tremblrel. 02, Created)		

FT	PRAM:	PF00595:	PDZ:	1.
FT	NON_TER	1	579	1
SO	SEQUENCE	579 AA:	65197 MW:	6FC3BF59 CRC32:
Query Match				
Best Local Similarity 32.8%; Score 351; DB 4; Length 579;				
Matches 60; Conservative 56; Mismatches 61; Indels 6; Gaps 5;				
D	60	EGVDEDEDEEE--DEIDRTDPL--Q--GRD--VRECELEKPADKDDDDLEGLEMMHLP	114	
Q	134	DAPPEEREEKEDDELQDTMLLSQNGPRDMMHMLIRKPGGQVDDLEIYEELHLC	193	
D	115	AFAAMTMSVRELCSVMLEFEVEEAGALLIEDGGELDSWYILNGTVEIS--HPDGKYNL	173	
Q	194	ALSHSTPVKRELAVLIPESHAKGGVYLFNGEGSTWYILKGSVNVVYIGKGVCTM	253	
D	174	FMNNGSGIIPITLDKQYMGIVRTKYDDCCQYPCIAODYKMLLNHYKNTNHYKEEGELYM	233	
Q	254	HSGDDGEGKLALVNDAPRAASIVLRKEDNCHELRVKKEDFNRIKDYEANTVRLKEDDDVL	313	
D	234	VHE 236		
Q	314	VLE 316		
RESULT: 13				
ID	014827	PRELIMINARY:	PRT:	1237 AA.
AC	014827:			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-MAY-1999 (TREMBLrel. 10, Last annotation update)			
DE	RAS-GRF2 (FRAGMENT).			
OS	Homo sapiens (Human).			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
CC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	TISSUE-BRAIN, HIPPOCAMPUS;			
RA	AMBOROH P.H.;			
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF023130; AAB80953.1. -			
DR	PRAM: PF00612; IO: 1.			
DR	PRAM: PF00169; PH: 2.			
DR	PRAM: PF00617; RASGERF, 1.			
DR	PRAM: PF00618; RASGERF, 1.			
DR	PRAM: PF00621; RHOGERF, 1.			
FT	NON_TER 1237 1237			
SO	SEQUENCE 1237 AA; 140848 MW; 37445EB4 CRC32:			
Query Match				
Best Local Similarity 26.9%; Score 326; DB 4; Length 1237;				
Matches 66; Conservative 72; Mismatches 100; Indels 7; Gaps 6;				
D	998	FESLSANAEALQITLLDHYIFRSIPYEEFLGOGMKLDKNERTPYIMKTSQHFNDSMLV	1057	
Q	606	FELMSSKDLAYQMTIYDMELEFCVHEHLEIYHTGRHNEFKTTANLDLFRFNEIDQPV	665	
D	1058	ASQIMNVAOVSSAANAIEKWNVAADICGLNNGVLEISALNBSAIVRLKTKTAYAKSK	1117	
Q	666	VTELCICLSQSKRVOLKKEFYIAAHCKEYNNLSFALYMGISNAYSKRLALTEKPLS	725	
D	1118	QTKALMDKLOKTVSSSEGRFKNLETLKNCNPPAVYLGWLTLDLAFIEEGTNPTEEGIV	1177	
Q	726	KFKFFVYEEFSLMDPSRNHRARIVLVALEPLLPFEMPLIKIMPTHTGNGTKTFYDN-LV	784	
D	1178	NFSKMRISHIIEIIEFOOFTSYRIDH-QP-KVAQYLLD--KDL-LIIDEQ-LTYELSLKI	1231	
Q	785	NFEKMRIMANTATATVYVYISQGFNPDAAQAKNKNQDVRSYRVLDANYIDQKRLTSQMSRL	844	
D	1232	EPRLP 1236		
Q	845	EPRLP 849		

ID	RESULT	14	PRELIMINARY:	PRT:	1189	AA.
TD	P70392					
AC	P70392					
DT	01-FEB-1997	(TREMblrel. 02, Created)				
DT	01-FEB-1997	(TREMblrel. 02, Last sequence update)				
DT	01-NOV-1999	(TREMblrel. 12, Last annotation update)				
DE	RAS PROTEIN-SPECIFIC GUANINE NUCLEOTIDE-RELEASING FACTOR 2					
DE	(RAS-GRF2).					
GN	RASGRF2.					
OS	Ras musculus (Mouse)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISUE-BRAIN;					
RC	MEDLINE: 94123458.					
RA	CHEN L., ZHANG L.J., GREER P., TUNG P.S., MORAN M.F.;					
RT	"A murine CDC25/ras-grf-related protein implicated in Ras					
RT	regulation.";					
RL	Dev. Genet. 14:339-346(1993).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	TISUE-BRAIN;					
RC	MEDLINE: 97184464.					
RA	FAN N.P., PAN W., ZHANG L., CHEN H., MORAN M.F.;					
RT	"Cloning and characterization of Ras-GRF2, a novel guanine nucleotide					
RT	exchange factor for Ras.";					
RL	Mol. Cell. Biol. 17:1396-1406(1997).					
DR	EMBL: U67326; AAC53058.1; -.					
DR	MGI:109137; RASgrf2.					
DR	PFAM: PF00612; IQ: 1.					
DR	PFAM: PF00169; PH: 2.					
DR	PFAM: PF00617; RASGEF: 1.					
DR	PFAM: PF00618; RASGEF: 1.					
DR	PFAM: PF00621; RHOGEF: 1.					
SO	SEQUENCE 1189 AA; 135250 MM; 35087683 CRC32;					
Query Match 5.1%; Score 320; DB 11; Length 1189;						
Best Local Similarity 26.9%; Pred. NO. 2,17e-38;						
Matches 66; Conservative 71; Mismatches 101; Indels 7; Gaps 6						
Db	950	PELISAAEELAEQTLTLDHIVPFSITPEEFLDQGMKIDKNERPPYIMKRSQHFNEKSNLY	1009			
QY	606	FELSSSDLAQMTQITDWELEFNCVHELTLYHTGGRNFEKFTTAMDLLRRENEIQFW	665			
Db	1010	ASQIMNADISSPNNAIEFWVAVADICQLNNGVLEIISALNRSPIYLKKTAKVSK	1069			
QY	666	VTEICLSQSLSKVQLLKKFKFIKAHCKEYKLNLSFPAIVMGSLNIAVSRLALVTEKLS	725			
Db	1070	QTKALMDKLOKTVSSGGRKKNRLRELKNCNPANPYIGMTITDPLAFLEGSTPRTFEGIV	1129			
QY	726	KEKFEVAFESLMDPSRNRHARLRVLVALEPLLPIMPELLIKMTPLTSGNFTFDN-LV	784			
Db	1130	NFSKRMISHIIRIIEISQFOQATYRIDQ-QP-KVIQYLYLD--KAL-VTDED-SLYELSLKI	1183			
QY	785	NFEKRMIAITARTATVTVYRSGCFNPDAAQNAKKNHDVRSYRQLVNIDNQRTLSQMSHRL	844			
Db	1184	EPRLP 1188				
QY	845	EPRLP 849				
RESULT	15	PRELIMINARY:	PRT:	557	AA.	
ID	015059					
AC	015059					
DT	01-JAN-1998	(TREMblrel. 05, Created)				
DT	01-JAN-1998	(TREMblrel. 05, Last sequence update)				
DT	01-NOV-1998	(TREMblrel. 08, Last annotation update)				
DE	K1AA0351.					
GN	Homo sapiens (Human).					

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 97349984.
 RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
 RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 4:141-150(1997).
 DR EMBL: AB02349; BAA20808.1; -
 DR PFAM: PF00617; RasGEF; 1.
 SQ SEQUENCE 557 AA; 62132 MW; 4CB4D1BA CRC32;

Query Match 4.98; Score 302; DB 4; Length 557;

Best Local Similarity 25.88; Pred. No. 9,10e-35;
 Matches 64; Conservative 66; Mismatches 108; Indels 10; Gaps 10;

Db 44 FDVLKVTPEEFASQITLMDIPVFKAIQPEELASCGWSKKEKHSIADPNVVAFTRRFNQVSF 103
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Qy 606 FE-L-MSSKRLAYQMTIYDWELENCVHELELIYHTEGHNFKKTTANLDLELRFNEIQF 663
 Db 104 WVVREI-LTAQTLKIRAEIISHFVKIATAKLELNTHSLMSVSAIQSAPIFRLTKTAL 162
 ||| ||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Qy 664 WVVTEICLSQ-LSKRVQLKKFKIKIAHCKEYKNLNSFPAIYVGLSNIASVRLALTWEK 722
 Db 163 LNRDKTTFEKLIDYLMKSEKDYKRTREYIRSLKMWPSIPYLGIVLIDLIIYDSAVPASGS 222
 | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Qy 723 LPSKFKKFTYAEFESLMDPSNNHRAIRLTVAKE-PLIPFMPLIKDMFTHEGNTFTD 781
 Db 223 IMENQORSNOMNIIIRIADLOVSCSYDLTTLPHVQKYLKSVR-YIEELQKFEVDNY- 280
 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Qy 782 NLVVFEEKRMRIANTAFVRYRSQ-PFNP-DA-AQANKNHQDVRSYVROLNVIDNQRTLS 838
 Db 281 KLSLRIEP 288
 : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Qy 839 QMSHRLRP 846

Search completed: Sat Apr 22 13:22:56 2000
 Job time : 128 secs.


```

OY 92 SRTQAGVMQVLLDEGVNLHVDGHHFQD-FYLFYRFIDDEHEDAPLPTFEKKCEDEEL 150
DB 528 SRSITLLSSVADPTLFLMTVSKPGEERSEPELEVYEEELTFKALSHSTMYKROLNSFV 587
OY 151 OOTMLLSQMGDAHMRMLRKPRGCRVDDLEIYEELIHLIKALSHSTMYKRELACVL 210
DB 588 KYEQVYHAGSVYEROGELIVYVYKAGVEYVNGK-IYCLLRGDDPGKALVNDLPR 646
OY 211 IRESNAKGGIVLFNOEGESTWITLKGSVNVYTGKGVNVTYIGKVGDDGKALVNDLPR 270
DB 647 AATVYTEDDSMFLVVDKHHFNQILHOEANTVRLKDYGEDVLEKVDYDIPGALNSN 706
OY 271 AASIVLRDNCHFLRVKDEDFMRILRDYEVANTVRLKEHDQVLYEKK--PAGNASVNG 328
DB 707 SCNFNGSVYVAGKAKKILEYVLEL-RIDALDDISELDEVEDITLHDAMPDNTVCN 765
OY 329 NSOPOOKYVMSGTPEKILIEHLETLRLATLNEATD-SVL-NDFIMHCVFMPTOLCP 386
DB 766 FLKSYFFPTVYATRDISTDCTEEVRCRRVVOEVYVWCSSLRVNFFL-NPTVNSFVEE 824
OY 387 ALVANHHAOPSGTEQEKMD-YALN-N-KRVIYRLVLOMAAY-GDLLOEDDVSMATLEE 442
DB 825 LFCHVLDKRRKLGEMDILFRIGSI-RSTRENMODVLARHPAIV--LDCGVLSAHTPCPV 881
OY 443 FYVYSDDARMIALAEQOLPELEKIVKQISEDAKAPQKKHVKVLLQOFMTGDERAQKQPI 502
DB 882 LPSDQNCNIIYIADTTCFLPRLRVDKTAELCELSRRKRSFAEPLNVEYKSGEKLIF 941
OY 503 RGSDEVLFVYVCMHTYTTIRVPVATSVKEVISAADAKIG-SGEGILLIKSSSGEKEVYL 561
DB 942 SPNDRAIPVLSLNSKLYVNVREERIPLVPMEDONGPTPSHSHSLIHLIDSOELAHOLEL 1001
OY 562 KRNDVSVEFTLTITNGELFCPRQEDSLPLPEDEGPYVGTGT-ELMSSDLAYQMTI 620
DB 1002 FHLQILRSTDSNELLXYOIGRESFPLSMFENLDLLVRRENVQHMSTTEILLATE-ENRM 1060
OY 621 YWELFNCVHELELYHYHFGRNHFKKTA-NLDLFLRRNEIOFWVWEICLSQSLSKRV 679
DB 1061 ELIKFKISATITAREKRDLLYFATTLGISHTSISRLILTMKSLPASPALKTFSELENLID 1120
OY 680 QLLKFKFIKAACHCKEKNLNSFALVMGSLNVAISRLATWEEKLPSKRRKKYAEESLMD 739
DB 1121 PPRNHMYRLVLYSKMSPYIPVPLILKDMFTHOGNKSFEVGLVNFPMHFAKIFRSF 1180
OY 740 PSNRHRAVLYIAKLEPPLIPMLIKDMTTHGCKNTFTDNLVNFEMKMIANTARTV 799
DB 1181 ROCKSQ-MDNGAHEFIEPQSLIRMLR 1206
OY 800 RYRSQPFNPDAQAQANKNHQDVRSYVR 826

```

```

CC 1- THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC 1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M82964; AAA34479.1; -.
DR PIR: PC1114; PC1114.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PRAM: PF00617; RASGEF; 1.
DR PRAM: PF00618; RASGEF; 1.
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;
KW Mitosis; Transmembrane.
FT NON_TER 1
FT TRANSMEM 1
SQ SEQUENCE 1095 AA; 125605 MM; 765862 CRC32;
Query Match 5.18; Score 317; DB 1; Length 1095;
Best local Similarity 32.5%; Pred. No. 1,63e-37;
Matches 67; Conservative 50; Mismatches 81; Indels 8; Gaps 6;
DB 808 LLDLSDIDYAKQLTIKHSFLFYKISPECDRTWGNKCYCMGSKNTEPISNHLTNV 867
OY 608 LMSSKDAY--QMTIYWELEFNCVHELELYHYHFGRNHFKKTA-NLDLFLRRNEIOFW 664
DB 868 VSFMIYVQTKRIKRILOFQFINVAACHLNNFSSLTALISALYSSPIRLKRTAAAY 927
OY 665 VYFICICLSQSKRVOLKKFKIAACHCKEKNLNSFALVMGSLNVAISRLATWEEKLP 724
DB 928 EYKLLLELNTLMDSAKNFIRYROLKISGDPCVPEFVYSLDTFTANGNPDLARN 967
OY 725 SKFKFYAEFESLMDPSNRHRAVR-LTVAKLEPPLIPFMPILIKDMTTHGCKNTFTD-N 782
DB 968 TVLVNFGKRVRLLELKEISVYQSSH 1013
OY 783 --LVNFEKMIANTARTVRY--RSQ 805

```

```

RESULT 2
ID CC25_SACKL STANDARD; PRT; 1095 AA.
AC Q02342;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CELL DIVISION CONTROL PROTEIN 25 (FRAGMENT).
GN CDC25.
OS Saccharomyces kluyveri (Yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92354938.
RA PRIGOSY T., GONZALES E., BROEK D.;
RT "Identification and analysis of a DNA fragment from Saccharomyces
RT kluyveri that can complement the loss of CDC25 function in
RT Saccharomyces cerevisiae."
RL Gene 117:67-72(1992).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,

```

```

RESULT 3
ID GMRP_RAT STANDARD; PRT; 1244 AA.
AC P28818;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GMRP) (P140 RAS-GRF).
GN RASGRF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN.
RX MEDLINE: 92350260.
RA SHOU C., FARUSWORTH C.L., NEEL B.G., FEIG L.A.;
RT "Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing
RT factor for Ras p21."
RL Nature 358:351-354(1992).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



```

OC Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90190870.
RA HUGHES D.A., FUKUI Y., YAMAMOTO M.;
RT "Homologous activators of ras in fission and budding yeast.";
RL Nature 344:355-357(1990).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN IS ESSENTIAL FOR MATING.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53254; CAA37345.1; -.
DR PIR: S28098; S28098.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00617; RASGEF; 1.
DR PFAM: PF00618; RASGEF; 1.
DR PFAM: PF00618; RASGEF; 1.
DR Guanine-nucleotide releasing factor; SH3 domain.
KW DOMAIN
FT SEQUENCE 911 AA: 105185 MW: 80580852 CRC32:
SQ
Query Match 4.5%; Score 278; DB 1; Length 911;
Best Local Similarity 22.2%; Pred. No. 9,016-30;
Matches 57; Conservative 75; Mismatches 119; Indels 6; Gaps 6;
DB 638 QLELDLS-LHNSPDPIYKDELIVLLPPREIAKQCLIEFQSFHSIRIQFLTKIMDLN 696
OY :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
584 EGFSDTLPLPEQEGPTVGVTGTFEELMSSKDLAYQMTIYDWELFNCVHELEIYHTFGHN 643
DB 697 RSPKRTSTELYSN-HLVNF-VTEIYOEBRRRTVLAFLYOVCYLLELNFAFLF 754
OY |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
644 -FKTTANLDLFRFPNETIQFWVTEICLSQSLSKRVOLKKFKIAHCKEYKNLSFF 702
DB 755 STISALNSSPIHRLKRTWANLNSKTLASFELINLTLEARKNFSNYRDLCEVCVPLF 814
OY :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
703 AIVMGISNIAVSRLALTWKELPSKFKFYAEFESLMDSNRHRAVRLVALLEPLPIFM 762
DB 815 GYFFDLDLFLKTKGNKDNFQNMINDKRTKVTIRILNEIKFQSGVGYMPNINEVOELLE 874
OY :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
763 PLLIKDMFTTHEGNKTFIDNLYNFEKMRMIANTARTVRYRSQFPNDAQANKNH-ODV 821
DB 875 ISRENTNNI-YORSLT 890
OY |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
822 KSYVOLNLYDNQRTLS 838
OY |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
RESULT 6 STANDARD: PRT: 1332 AA.
ID SOS2_HUMAN
AC 007890; 015503;
DB 15-JUL-1999 (Rel. 38, Created)
DB 15-JUL-1999 (Rel. 38, Last sequence update)
DB 15-JUL-1999 (Rel. 38, Last annotation update)
DE SON OF SEVENLESS PROTEIN HOMOLOG 2 (SOS-2).
GN SOS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93262494.
RA CHARDIN P., CAMONIS J.H., GALE N.W., VAN AELST L., WIGLER M.H.,
RA BAR-SAGI D.;

```

```

RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
RT to GRB2.";
RL Science 260:1338-1343(1993).
RN [2]
RP SEQUENCE OF 724-1296 FROM N.A.
RX TISSUE-PLACENTA;
RA FATH I., APIOU F., DUTRILLAUX B., TOCQUE B.;
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L13858; AAA35914.1; -.
DR EMBL: L20686; AAA91852.1; -.
DR MIM: 601247; -.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00617; RASGEF; 1.
DR PFAM: PF00618; RASGEF; 1.
DR PFAM: PF00621; RhoGEF; 1.
DR Guanine-nucleotide releasing factor.
KW DH.
FT DOMAIN 200 441
FT DOMAIN 442 546
FT DOMAIN 546 561
FT DOMAIN 561 575
FT DOMAIN 575 758
FT DOMAIN 758 1180
FT DOMAIN 1180 1208
FT DOMAIN 1208 1278
FT CONFLICT 778 778
FT CONFLICT 861 861
FT CONFLICT 861 861
FT CONFLICT 948 948
FT CONFLICT 948 948
FT CONFLICT 999 999
FT CONFLICT 1032 1032
FT CONFLICT 1032 1032
FT CONFLICT 1042 1042
FT CONFLICT 1112 1114
FT CONFLICT 1159 1159
FT CONFLICT 1296 1296
SQ SEQUENCE 1332 AA: 153030 MW: 96708378 CRC32:
Query Match 4.4%; Score 272; DB 1; Length 1332;
Best Local Similarity 25.8%; Pred. No. 1,336-28;
Matches 59; Conservative 70; Mismatches 90; Indels 10; Gaps 7;
DB 752 FESPPPIEWHISKPGQFETDMLDPIEIAQRLLESIDLYRKVQPSLVGSVTKED 811
OY :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
586 FDSITPLPEQEGPTVGVTGTFEELMS--SKDLAYQMTIYDWELFNCVHELEIYHTFGHN 643
DB 812 KEINSPILKMIHNTNLTLEFEKCIYEAEFEEERVAVLSRIEILQVQDINFGVLE 871
OY :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
644 FKTTANLDLFRFPNETIQFWVTEICLSQSLSKRVOLKKFKIAHCKEYKNLSFFA 703
DB 872 IYSAVNSVSYRRDHTFEALQERKRKL-D-EA-VELSDHEFKYLVKLSINPCVPFF 928
OY 704 IYMGISNIAVSRLALTWKELPSKFKFYAEFESLMDSNRH-RAVLLYVAKLEPLPIFM 762
DB 929 GIVLTNLTKEGNDFLKRGKDLINFSKRRVAETGEIQOYQNPY 977
OY :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
763 PLLIKDMFTTHEGNKTFI--D-NLVNFEKMRMIANTARTVRYRSQFP 807
OY |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
RESULT 7 STANDARD: PRT: 1297 AA.
ID SOS2_MOUSE
AC 002384;

```

```
DR DT 15-JUL-1999 (Rel. 38, Created)
DT DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE DE 15-JUL-1999 (Rel. 38, Last annotation update)
GN GN SON OF SEVENTEENLESS PROTEIN HOMOLOG 1 (SOS-1) (MSOS-1).
    SOS1.
OG Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
RN Rn Eurotyla; Rodentia; Sciurognathi; Muridae; Mus.
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN-SWISS: TISSUE-EYE;
RA MEDLINE; 92335328.
RT BOWTELL D., FU P., SIMON M., SENIOR P.;
RL "Identification of murine homologues of the Drosophila son of
[2] seventeenless gene: potential activators of ras.";
R proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
RP STRUCTURE BY NMR OF 415-548.
RA KOSHIBA S., KIGAWA T., KIM J.-H., SHIROZU M., BOWTELL D., YOKOYAMA S.;
RX MEDLINE; 97360234.
RT "The solution structure of the pleckstrin homology domain of mouse
J. Mol. Biol. 269:579-591(1997)."
CC -I FUNCTION: PROMOTES THE EXCHANGE OF GTP-BINDING GDP FOR GTP (BY
SIMILARITY);
CC -I TISSUE SPECIFICITY: EXPRESSED IN MOST EMBRYONIC AND ADULT TISSUES.
CC -I SIMILARITY: CONTAINS 1 DEL-HOMOLOG DOMAIN (DH).
CC -I SIMILARITY: CONTAINS 1 PH-DOMAIN.
CC -I SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its contents remain unchanged
and modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR DR EMBL; Z11574; CAA77662.1; -.
DR EMBL; Z11578; CAA77665.1; -.
DR PDB; 1PWS; 15-MAY-97.
DR MGD; MG1:98354; SOS1.
DR PROSITE; PS00720; GDS.CDC25; 1.
DR PROSITE; PSS0003; PH.DOMAIN; 1.
DR PFAM; PF00169; PH; 1..
DR PFAM; PF00617; RasGef; 1.
DR PFAM; PF00618; RasGen; 1.
DR PFAM; PF00621; RhogEF; 1.
KW Guanine-nucleotide releasing factor; 3D-structure.
FT DOMAIN 202 443 DH.
FT FT 444 548 PH.
FT DOMAIN 777 963 RASGEF.
FT DOMAIN 1244 1247 POLY-PRO.
SQ SEQUENCE 1319 AA; 150882 MW; A317FA3B CRC32;

Query Match      4.1%   Score 255; DB 1; Length 1319;
Best Local Similarity 25.4%; Pred.No. 2.5Se-25;
Matches 58; Conservative 64; Mismatches 98; Indels 8; Gaps 6;
```


QY 671 LC-SQSKRVQLKFKFIKIAHCKEYKLNLSFPAIVMGLSNIAVSRLATMKEKPSKFKK 729

Db 1189 NIKNMKNKLNSSRNENEYR-DVTKFGSEPCYPPFEGVITSDITFVYHGNPDLYNKTQV 1247

QY 730 FYAEFSLMDPSRNHAFYRLTYAK-LEP-PLIPFMDLKIDMTFHEGNKFTI-D-N-LV 784

Db 1248 NFAKRAKTSRIVSGIDRFKTTGYNFOEVEPEIQK 1280

QY 785 NFEKRMIMANTARTVRYRYSQPPN-PDAQAQAK 816

RESULT 11

ID CC25_YEAST STANDARD; PRT; 1589 AA.

AC P04821;

DT 13-AUG-1987 (Rel. 05, Created)

DT 01-JAN-1988 (Rel. 06, last sequence update)

DT 15-JUL-1999 (Rel. 38, last annotation update)

DE CELL DIVISION CONTROL PROTEIN 25.

GN CDC25 OR CTN1 OR YLR310C OR L2142.6.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 87131091.

RA BROEK D., TODA T., MICHAELI T., LEVIN L., BIRCHMEIER C., ZOLLER M., POWERS S., WIGLER M.;

RT "The S. cerevisiae CDC25 gene product regulates the RAS/adenylylate cyclase pathway.";

RL Cell 48:789-799(1987).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 86220116.

RA CAMONIS J.H., KALKRINE M., GONDRE B., GARREAU H., BOY-MARCOITE E., JACQUET M.;

RT "Characterization, cloning and sequence analysis of the CDC25 gene which controls the cyclic AMP level of Saccharomyces cerevisiae.";

RL EMBO J. 5:375-380(1986).

RN [3]

RP SEQUENCE FROM N.A.

RX STRAIN-8288C / AB972;

RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z., FAVELLO A., ELDON L., GATTUNG S., GRECO T., KIRSTEN J., KOCACA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D., JOHNSON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENZES S., MILLER N., NHAN M., PAULEY A., PELUSO D., RIEKEN L., RILES L., TATCH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;

RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.

RN [4]

RP DOMAINS.

RX MEDLINE: 89181526.

RA MONDER T., MINK M., KUNTZEL H.;

RT "Domains of the Saccharomyces cerevisiae CDC25 gene controlling mitosis and meiosis.";

RL Mol. Gen. Genet. 214:271-277(1988).

RN [5]

RP FUNCTION.

RX MEDLINE: 91203884.

RA JONES S., VIGNAIS M.L., BROACH J.R.;

RT "The CDC25 protein of Saccharomyces cerevisiae promotes exchange of guanine nucleotides bound to ras.";

RL Mol. Cell. Biol. 11:2641-2646(1991).

CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START, THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.

CC -----

CC THIS SMIS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X03579; CAA27259.1; -

DR EMBL: M15458; AAA34478.1; -

DR EMBL: U17247; AAB67360.1; -

DR EMBL: U20618; AAB64528.1; -

DR PIR: A26596; RGRYC5.

DR HSSP: P29355; ISEM.

DR SGD: L0000263; CDC25.

DR PROSITE: PS00720; GDS_CDC25; 1.

DR PROSITE: PS00002; SH3; 1.

DR PFAM: PF00018; SH3; 1.

DR PFAM: PF00617; RASGEF; 1.

DR PFAM: PF00618; RASGEF; 1.

KM Guanine-nucleotide releasing factor; Cell division; Cell cycle;

KW Mitosis; Transmembrane; SH3 domain.

FT TRANSMEM 1452 1473 POTENTIAL.

FT DOMAIN 58 128 SH3.

FT CONFLICT 497 497 I -> Y (IN REF. 2).

FT CONFLICT 954 963 PYGHHEPKN -> LSVYIMNLSR (IN REF. 2).

SQ SEQUENCE 1589 AA; 179091 MM; A28FA88A CRC32;

Query Match 4.18; Score 255; DB 1; Length 1589;

Best Local Similarity 25.58; Pred. No. 2.55e-25;

Matches 71; Conservative 77; Mismatches 117; Indels 13; Gaps 10;

Db 1228 VSEKFGAEDLLQKINEKINEKEFPVKKQDSVSNVQTKRDNKSP1-HMS5-SSL 1285

QY 537 VADKLGSGEGLI--IV-KMSGGEKVLKPNV-SVFTTLTINGSLFACPREQDSLPL 592

Db 1286 PSSASSAFPRKRLKLLDPTTYATQTLVLEHDIYLRITPECLDRMGKRYCNMGSP 1345

QY 593 PDEGPGVGTGTFELMSSKDLAY--QMTIYDWELFNCVHELELIYHFGHNRKT-TA 649

Db 1346 NIKETLANANTLNFESHTYKADYKTRSKLTQYFVYVAOCHKELNFSMTAIVSALY 1405

QY 650 NLDLFLRPFNEIOPWVYVTEICLSQSKRVOLKFKFIKIAHCKEYKLNLSFPAIVMGLS 709

Db 1406 SSPYRLKKTWDIVSTESDLKLNLMDSKRNFKYKTRBELRSYTDVACYPFEGVYISD 1465

QY 710 NIVASRLATLWELKPSKFKFYAEFESLMDPSRNRAVR-LTVAKLEPPLIPFMDLKID 768

Db 1466 LTFEPWGNPEFLNSNININFSKRTKIANIVEIISFK 1503

QY 769 MFTTHEGNKFTI-D--NLVNFEMKIMANTARTVRYR 803

RESULT 12

ID SOS_DROME STANDARD; PRT; 1595 AA.

AC P26675;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, last sequence update)

DT 15-JUL-1999 (Rel. 38, last annotation update)

DE SON OF SEVENTEEN PROTEIN.

GN SOS.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-OREGON-R;

RX MEDLINE: 92141820.

RA BONFINI L., KARLOVICH C.A., DASGUPTA C., BANERJEE U.;

RT "The Son of seventeen gene product: a putative activator of Ras.";

RL Science 255:603-606(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92034591.

RA SIMON M.A., BOWTELL D.D.L., DODSON G.S., LAVERY T.R., RUBIN G.M.;

RT "Ras1 and a putative guanine nucleotide exchange factor perform

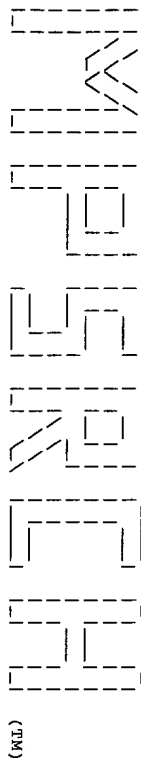
[illegible]

FT CONFLICT 214 214 T -> I (IN REF. 1).
FT CONFLICT 260 260 A -> V (IN REF. 1).
SQ SEQUENCE 376 AA; 42609 MM; B4F7833B CRC32;

Query Match 3.4%; Score 209; DB 1; Length 376;
Best Local Similarity 29.7%; Pred. No. 9.48e-17;
Matches 33; Conservative 31; Mismatches 44; Indels 3; Gaps 3;

Db 133 FAHLEDEQKTMIDAM-FPVEKSAGETIIEGEGEDNFYIDKGTVDYVYNHEVYL-TIN 190
Qy 195 LSHLSTVKKRELAVGLIFESHAKGQTVLENOEGESTWYIILKGSVNVVIYGRGVCTLH 254
Db 191 EGGSPGELALIVGPRAATVIKTDV-KLMAIDRLTYRIILMGSVTKKKRM 240
Qy 255 EGDDEGKLALVNDAPRAASIVLRDNCHEFLRVDKEDENRIILDYEAINTVRL 305

Search completed: Sat Apr 22 13:20:31 2000
Job time : 54 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat Apr 22 13:18:40 2000; MasPar time 36.15 Seconds
Tabular output not generated. 941.038 Million cell updates/sec

Title: >US-09-422-999-18
Description: (1-849) from US09422999.pep
Perfect Score: 6222
Sequence: 1 MVLKRALTMGTGFLTRFTHL.....VIDNQRTLSQMSHRLPPRRP 849

Scoring table:
PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 53.983; Variance 117.292; scale 0.460

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1922	30.9	1213	2	S42368	guanine nucleotide re	0.00e+00
2	317	5.1	1095	2	PC1114	SKDC25 protein - yea	6.73e-31
3	310	5.0	1244	2	S29083	guanine-nucleotide-re	9.95e-30
4	291	4.7	1260	2	S28407	guanine nucleotide-ex	1.40e-26
5	287	4.6	1275	2	A38985	nucleotide exchange f	6.36e-26
6	278	4.5	911	2	S28098	ste6 protein - f1sio	1.89e-24
7	266	4.3	1297	2	S25714	son-of-sevenless-2 pr	1.67e-22
8	255	4.1	1333	2	A37488	ras guanine nucleotid	9.72e-21
9	253	4.1	1333	2	S30356	CDC25 protein homolo	2.03e-20
10	255	4.1	1336	2	S25716	Ras guanine nucleotid	9.72e-21
11	255	4.1	1589	1	RGBYC5	cell division control	9.72e-21
12	251	4.0	1596	2	A41216	guanine nucleotide ex	4.22e-20
13	240	3.9	1048	2	S64758	SCD25 protein (versio	2.32e-18
14	236	3.8	1250	2	S14177	SCD25 protein (versio	9.82e-18
15	216	3.5	378	1	OKGARI	guanine kinase (EC 2.	1.21e-14
16	210	3.4	1435	1	BVBV1L	guanine nucleotide-re	9.89e-14
17	208	3.3	376	1	OKKMR	protein kinase (EC 2.	1.98e-13
18	205	3.3	377	1	OKKMR	protein kinase (EC 2.	5.60e-13
19	207	3.3	381	1	A60669	protein kinase (EC 2.	2.80e-13
20	204	3.3	381	1	OKMSR1	protein kinase (EC 2.	7.91e-13
21	206	3.3	411	2	S18634	cgsl protein - f1sio	3.96e-13
22	203	3.3	416	1	OKBYRC	protein kinase (EC 2.	1.12e-12
23	201	3.2	379	1	OKHUR1	protein kinase (EC 2.	2.22e-12

RESULT ENTRY TITLE	1	ALIGNMENTS
ORGANISM	542368	#type complete
DATE	07-Oct-1994	guanine nucleotide releasing factor homolog - Caenorhabditis elegans
ACCESSIONS	07-Aug-1998	#formal_name Caenorhabditis elegans
REFERENCE	S42368	#submission
#authors	Smith, A.	submitted to the EMBL Data Library, March 1994
#accession	S42368	preliminary
#status	1-1213	#label SMI
#molecule_type	DNA	#residues
#cross-references	EMBL:230423; NID:g458479; PID:g458480	
GENETICS	196/3; 238/1; 263/3; 312/3; 444/3; 519/3; 550/2; 669/2;	
#introns	742/1; 776/3; 1072/2; 1102/3; 1107/3	
CLASSIFICATION	#superfamily cAMP receptor protein cyclic nucleotide-binding domain homology; CDC25-type guanine nucleotide exchange activator homology	
FEATURE	574-689	#domain cAMP receptor protein cyclic nucleotide-binding
987-1206	#domain homology #label CA1\	
SUMMARY	#length 1213	#molecular-weight 139379 #checksum 3211
Query Match	30.9%;	Score 1922; DB 2; Length 1213;
Best Local Similarity	38.5%;	Pred. No. 0.00e+00;
Matches	311; Conservative	204; Mismatches 259; Indels 33; Gaps 24;

Db	414	EQIIRDSGILHRKMLTDHNOVIRDTTEHTRVQCMGAEIMDLTLFVSTGSSSL	473
Qy	39	EKLIR-AGKILNALISRAPIHMRKYLKTYRCCVGTLEVDMM-----ID-ETPCVN	91
Db	474	SRIOKSAIWOVLNGLSHIDGEOFLDKTNSYRW---OQ---FSSRNKVASIEV	527
Qy	92	SRTQAVGMVQVLLDEGVNLHVDQEHFOD-FYLFYRFIDDEHEDAPLPTEEKKCEDDEL	150
Db	528	SKSTPLSSVAPETFLFMVSKSGFERSPELEVVYEELTFKALSHSTWKRQLSNFV	587
Qy	151	QDTMLLSQMGPDAMHMRILKRPQGRIVYDDELTTEELHFKALSHSTVYKRLAGVL	210
Db	588	KVEQYVHAGSVYFROGELGVYVIYVLCGAVEVNVGK-IYCLLRGDEFGKLTAVNDPR	646

QY 211 IFESHAKGTVLFGNOEGESTSWYIILKGSVNVVYIGKGVCTLHREGDDEFGKILATLVNDAPR 270
DB 647 AATVTEEDDSMEFLVYDKHFNQIILHOEVANTVRLKDGEDLVLEKVIDIRGALLENSSN 706
QY 271 AASIVLRDNCFPLRVYDKDFENKILRDVANTVRLKEDHODVLEKY--PAGNRASNOG 328
DB 707 SCNFNGSVMAKAEKILEVLET-RIDALGDDISELDFEDEDILTTHAFMDPNVTCN 765
QY 329 NSQPOOKYVMSGTEPKLEHLEFTRILEATLENATD-SVL-NDPIMHCVFMPRTQJCP 386
DB 766 FLKSYIFRPIRYATDSTIDSTEVEYRCKRVYOVYVWCSLLRNFL-NDPTNSVEE 824
QY 387 ALVAHHAQPSQGTQEKMD-YALN-N-KRVYRLVLQWAMV-GDLQEDDVSAFLEE 442
DB 825 LFCYHIDDKRLGEMDILTRIGST-RSTREMOQLVARHPATV-LDQVLSAHTPCPV 881
QY 443 FYVSYSDDAKMTALKEOLPELEKTVKQISEDAKAPQKHVYLQDFMTGDERAKKROP 502
DB 882 LPSDYCNQIILVADTYTCVFLPIRVDTAEICELSRRMSEFAEPLNVEYKNGEKLIF 941
QY 503 RGSDEVLFKVCMDHTYTTIRVPVATSVKEVISAVALKIG-SGEGILLIVKMSGGEKAYVL 561
DB 942 SPNDRAIPVLSLNSKLYVNRBEIPLVPMEDQNPSPSSHSILHLIDSOELAHQLFL 1001
QY 562 KENDVSVEFTLTINGRLEFACPREQFDSLPLPEDEGPTVGVGF-ELMSSKDLAYQMTI 620
DB 1002 FHLQILRSDSNELLVOYGRSESPLSMPENDLLVRRENVQHSSTELLATE-ENRM 1060
QY 621 YMELEPCVHELELYHFFGRHNEKKTVA-NLDFLRNFNELOFVWVEICCSOLSKRV 679
DB 1061 ELTKRFISATTAAREYRDLTFVATTLGSHTSISRLTLWMSKLPSPALKTFSLELND 1120
QY 680 QLKFKFIKAACHKEKKNNSFFALVMGSLNVAIRLALTWKLPSPKRTKTYAESLMD 739
DB 1121 PPRNHRMRLVSKMSSPIFPVPLIKDMFTHOGKNSFTNGVNEFKMHAKIFRSF 1180
QY 740 PPRNHRMRLVSKMSSPIFPVPLIKDMFTHOGKNSFTNGVNEFKMHAKIFRSF 1180
DB 1181 ROCKSO-MDNGAEHEFIEPOSTLRMLR 1206
QY 800 RYRROQFNPDAQAQKNHODVRSYR 826

RESULT 2
ENTRY PC1114 #type fragment
TITLE SKDC25 protein - yeast (Saccharomyces kluyveri) (fragment)
ORGANISM #formal_name Saccharomyces kluyveri
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997
ACCESSIONS PC1114
REFERENCE PC1114
#authors Priegozy, T.; Gonzales, E.; Broek, D.
#journal Gene (1992) 117:67-72
#title Identification and analysis of a DNA fragment from Saccharomyces kluyveri that can complement the loss of CDC25 function in Saccharomyces cerevisiae.
#cross-references MVID:92354938
#accession PC1114
#molecule_type DNA
#residues 1-1095 #label PRI
#cross-references GB:M82964; NID:g171186; PID:g171187
GENETICS
#gene SKDC25
CLASSIFICATION #superfamily CDC25-type guanine nucleotide exchange activator
KEYWORDS homology
FEATURE transmembrane protein
808-1049 #domain CDC25-type guanine nucleotide exchange activator
SUMMARY homology #label SOS
#length 1095 #checksum 2353

Query Match 5.1%; Score 317; DB 2; Length 1095;
Best Local Similarity 32.5%; Pred. No. 6.73e-31;

Matches 67; Conservative 50; Mismatches 81; Indels 8; Gaps 6;
DB 808 LLDLSDIDYAKQLTIKHSFLFKISPECDLRWGNKCYCMGSKNITEPISNHLTVN 867
QY 608 LMSKDLAY--QMTIYDWELFNCVHELELYHFFGRHNEKKTVA-NLDFLRNFNIQFW 664
DB 868 VSFMTIVQTDIKRIQIOLPFIWVAACHLNFSSITAFISALYSSPIRLKRTAAVP 927
QY 665 VMEICCSQSLSKRVQLKFKFIKAACHKEKKNNSFFALVMGSLNVAIRLALTWKLP 724
DB 928 EYKKLLEINTLMDSAKNFIRYRLQKLSIGDEPCVPEGVVSLDTFTANGNDFLHRN 987
QY 725 SKFKFYAEFESLMDPSRNHRAVR-LTVAKLEPLILPFMPLIKDMFTHEGKTEID-N 782
DB 988 TVLVNFKRVRILEILKEISVYQRSH 1013
QY 783 --LVNFKRMRIANTARTVRY--RSQ 805

RESULT 3
ENTRY S29083 #type complete
TITLE guanine-nucleotide-releasing protein - rat
ALTERNATE_NAMES CDC25 protein homology
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999
ACCESSIONS S29083
REFERENCE S29083
#authors Shou, C.; Farnsworth, C.L.; Neel, B.G.; Fely, L.A.
#journal Nature (1992) 358:351-354
#title Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor for Ras p21.
#cross-references MVID:92350260
#accession S29083
#molecule_type mRNA
#residues 1-1244 #label SHO
#cross-references EMBL:X67241; NID:957664; PID:957665
#note the authors translated the codon GAG for residues 135 and 137 as Gln
CLASSIFICATION #superfamily CDC25-type guanine nucleotide exchange activator
homology: CDC24 homology; pleckstrin repeat homology
FEATURE 240-426
1005-1241 #domain CDC24 homology #label CD24
SUMMARY homology #label SOS
#length 1244 #molecular-weight 14266 #checksum 2921

Query Match 5.0%; Score 310; DB 2; Length 1244;
Best Local Similarity 26.4%; Pred. No. 9.95e-30;
Matches 65; Conservative 74; Mismatches 98; Indels 9; Gaps 7;
DB 1005 FENHSAEIALDQLTLDHLVFKSIPYEEFGQGMKADKNERPIYIMKTRHNSNLI 1064
QY 606 FELMSSDLDAYQMTIYDWELFNCVHELELYHFFGRHNEKKTVA-NLDFLRNFNIQFW 665
DB 1065 ASEL-LRNEVSARASTIEKVAVADICRLHYNVAVLETTSSINSAIFRLKRTWLKVS 1123
QY 666 VTEICLCSQ--LSKRVOILKFKFIKAACHKEKKNNSFFALVMGSLNVAIRLALTWKLP 724
DB 1124 KQTKSLFDKLOKLVSVDGRKRLRETLRNCDDPCVPLVGLYTLDAFLFEGTPNVTEDQL 1183
QY 725 SKFKFYAEFESLMDPSRNHRAVR-LTVAKLEPLILPFMPLIKDMFTHEGKTEID-N 783
DB 1184 VNSKRMAMSHIIREROQOITYKLEP-QP-KVYQL--VDETVLDDF-SLEASLR 1237
QY 784 VNEFKRMRIANTARTVRYRSDPFNDAAQANKHODVRSYVQLVINDORTLSQMSHR 843
DB 1238 IEPKLP 1243
QY 844 LEPRRP 849

RESULT 4

	RESULT	5
	ENTRY	A38985
	TITLE	nucleotide exchange factor CDC25 - human
	ALTERNATE_NAMES	Ras-specific guanine nucleotide-releasing factor
	ORGANISM	#formal name Homo sapiens #common name man
	DATE	19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 10-Sep-1997
	ACCESSIONS	A38985; A46199; I58371
	REFERENCE	A38985 #authors Wei, W.; Broek, D. #submission Submitted to GenBank, December 1994 #description Cloning and analysis of the full length human cdc25 cDNA, a ras-specific nucleotide exchange factor.
	#accession A38985	translated from GB/EMBL/DBJ
	##status ##molecule_type mRNA	
	##residues 1-1275	##label WEI
	##cross-references GB:I26584; NID:g433719; PID:g433720	
	REFERENCE A46199	
	#authors Wei, W.; Mosteller, R.D.; Sanyal, P.; Gonzales, E.; McKinney, D.; Dasgupta, C.; Li, F.; Liu, B.X.; Broek, D.	
	#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:7100-7104	
	#title Identification of a mammalian gene structurally and functionally related to the CDC25 gene of Saccharomyces cerevisiae.	
	##cross-references MIMD:92357779	
	#accession A46199	
	##molecule_type mRNA	
	##residues 1047-1054,'A',1056-1112,'G',1114-1135,'C',1137-1275	
	##experimental_source Brain	
	##note sequence extracted from NCBI backbone (NCBIN:111098, NCBIP:111099)	
	REFERENCE I58371	
	#authors Schlegeloffer, F.; Faure, M.; Rath, I.; Chevallier-Multon, M.C.; Apicou, F.; Dutrillaux, B.; Sturani, E.; Jacquet, M.; Toque, B.	
	#journal Oncogene (1993) 8:1477-1485	
	#title Identification of a human guanine nucleotide-releasing factor (H-GRF5) specific for Ras proteins.	
	##cross-references MIMD:93275641	
	#accession I58371	
	##status preliminary; translated from GB/EMBL/DBJ	
	##molecule_type mRNA	
	##residues 787-1275	##label RES
	##cross-references GB:S62035; NID:g386046; PID:g386047	
	CLASSIFICATION #superfamily CDC25-type guanine nucleotide exchange activator homology; CDC24 homology; plexstrin repeat homology	
	FEATURE	
	240-426	#domain CDC24 homology #label CD24\
	1036-1272	#domain CDC25-type guanine nucleotide exchange activator homology #label SOS
	SUMMARY	#length 1275 #molecular_weight 145381 #checksum 4961
	Query Match	4.6%; Score 287; DB 2; Length 1275;
	Best Local Similarity	25.6%; Pred. No. 6,36e-26;
	Matches	40; Conservative 49; Mismatches 66; Indels 1; Gaps 1;
	Db	1086 KHPNDISNLASLEIRNEDINAVSAIEKVVAVADICRLHYNAVLETSSMNSRAIFR 1145
	QY	656 RRPEWIOFWVTTEICLOSLSSTKVLKKFRIKAIACEKYNLNFPAIVGLSMIAVSR 715
	Db	1146 LKKTWLKVSKOTAFALDDLOKLTVSSSGRRKNLRALKNKDPCPVYILGMVITLDLAFTFG 1205
	OY	716 LALTWEALPSEKFKFAEFESLMDPSRNRAATRYAKLEPPLPFIPLLIKDMFTHEG 775
	Db	1206 TPNTYEDGLNVFSKMNIISHIREIROFOOTAAYKIE 1241
	OY	776 NKTFIDN-LVNFEKMRVIANTARTAVTAYSQDPFPD 810
	RESULT	6
	ENTRY	S28098
		#type complete

```

TITLE      ste6 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM   #formal_name Schizosaccharomyces pombe
DATE       17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
09-Sep-1997

ACCESSIONS
REFERENCE   S28098
#authors    Hughes, D.A.; Fukui, Y.; Yamamoto, M.
#journal     Nature (1990) 344:355-357
#title       Homologous activators of ras in fission and budding yeast.
#cross-references MUID:90190870
#accession   S28098
#molecule_type DNA
#residues    1-911 #label HUG
#cross-references EMBL:X53254; NID:g5100; PID:g5101

GENETICS
#gene        ste6
#superfamily CDC25-type guanine nucleotide exchange activator

FEATURE
659-894      #domain CDC25-type guanine nucleotide exchange activator
              #homology #label SOS

SUMMARY      #length 911 #molecular_weight 105185 #checksum 2084

Query Match      4.5%; Score 278; DB 2; Length 911;
Best Local Similarity 22.2%; Pred. No. 1,896-24;
Matches 57; Conservative 75; Mismatches 119; Indels 6; Gaps 6;

Db 638 OLEEDLS-LHNSPDIYIKDELVLPPREIAKOLCIEPFSHISRIQFTKIMDLN 696
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 584 EGFDSLTPLEDEGPTVGTGTFELMSKDLAYQTTITDWEIPLFCVHELELITFTGRRN 643
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 697 RESPEKSTFYLNS-HLVNF-VETETIYQEEPRRRNVLAFTQVCDYLRRLNPFSLF 754
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 644 -FKKTANLDELFRFNEIQFWYVEICLSQSLSKRVQLKKFIKIAHCKEYKNLSFF 702
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 755 STISLANSPIHRLKRYANLNSKTLASFELNNLTAKRNKFSNYRDCLECVLPVPL 814
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 703 AIVMGLSNIAVSRLALTWELKPSKFKFYAFESLMPSRNHRAYRLVATLEPPLIPFM 762
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 815 GYVFETDLFELKTKGNKNQNNIIPDKRTKVTRLINEIKKPSQVGMNPINVEQELNVEY 874
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 763 PLLIKDMFTFHEGKNTFIDNLVNEFKRMNANTARTVATYRSQPNPDAAQANKNH-ODV 821
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 875 ISREKRTNNI-YQRSLLT 890
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 822 RSYVROLNVIDNQRTLS 838
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT      7
ENTRY       S25714 #type fragment
TITLE       son-of-sevenless-2 protein - mouse (fragment)
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
17-Mar-1999

ACCESSIONS
REFERENCE   S25714
#authors    Bowtell, D.; Fu, P.; Simon, M.; Senior, P.
#journal     Proc. Natl. Acad. Sci. U.S.A. (1992) 89:6511-6515
#title       Identification of murine homologues of the Drosophila Son of
              sevenless gene: potential activators of ras.
#cross-references MUID:92335328
#accession   S25714
#status      preliminary
#molecule_type mRNA
#residues    1-1297 #label BOW
#cross-references EMBL:Z11664; NID:g54136; PID:g54137
CLASSIFICATION
#superfamily CDC25-type guanine nucleotide exchange activator
#homology; pleckstrin repeat homology

FEATURE
405-507      #domain pleckstrin repeat homology #label PLK\
739-982      #domain CDC25-type guanine nucleotide exchange activator
              #homology #label SOS

SUMMARY      #length 1297 #checksum 7302

```

```

Query Match      4.3%; Score 266; DB 2; Length 1297;
Best Local Similarity 24.9%; Pred. No. 1,676-22;
Matches 57; Conservative 70; Mismatches 92; Indels 10; Gaps 7;

Db 717 FESSPPVEWHISRGQFETFDLMTLHPLEIAQTLLESIDLYRAVQPSLGVSWTKED 776
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 586 FDSLTPLEDEGPTVGTGTFELMS--SKDLAYQMTIYDWEIPLFCVHELELITFTGRRN 643
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 777 KEINSPRLKMHRTNLTLMFEKCIVENEFERAVLSRIEILQVQDINNENGYLE 836
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 644 FKTTANLDELFRFNEIQFWYVEICLSQSLSKRVQLKKFIKIAHCKEYKNLSFFA 703
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 837 IVSAVNSVSYRLDHTFEALQER-KRRILD-DA-VELSDHFEKKYLVKLKSNPCVPFF 893
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 704 IVMGLSNIAVSRLALTWELKPSKFKFYAFESLMPSRNH-RAYRLVATLEPPLIPFM 762
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 894 GIYLTNLTKEGNSDFLKRKGDILNFSKRRVAETIGEIQYQONPY 942
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 763 PLLIKDMFTFHEGKNTFI--D-NLVNFEKRMNANTARTVATYRSQPF 807
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT      8
ENTRY       A37488 #type complete
TITLE       Ras guanine nucleotide exchange factor son-of-sevenless (sos)
ORGANISM    1 - human
DATE        03-Mar-1994 #sequence_revision 18-Nov-1994 #text_change
18-Mar-1997

ACCESSIONS
REFERENCE   A37488
#authors    Chardin, P.; Camonis, J.H.; Gale, N.W.; van Aelst, L.;
              Schlessinger, J.; Wigler, M.H.; Bar-Sagi, D.
#journal     Science (1993) 260:1338-1343
#title       Human Sos1: a guanine nucleotide exchange factor for Ras that
              binds to GNB2.
#cross-references MUID:93262494
#accession   A37488
#status      preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues    1-1333 #label CHA
#note        sequence extracted from NCBI backbone (NCBIP:132148)
CLASSIFICATION
#superfamily CDC25-type guanine nucleotide exchange activator
#homology; pleckstrin repeat homology

FEATURE
442-544      #domain pleckstrin repeat homology #label PLK\
776-1019      #domain CDC25-type guanine nucleotide exchange activator
              #homology #label SOS

SUMMARY      #length 1333 #molecular_weight 152473 #checksum 7142

Query Match      4.1%; Score 255; DB 2; Length 1333;
Best Local Similarity 25.4%; Pred. No. 9,726-21;
Matches 58; Conservative 65; Mismatches 97; Indels 8; Gaps 6;

Db 754 FQSSPPTVEWHISRGFIETFDLTLHPLEIAQTLLESIDLYRAVQPSLGVSWTKED 813
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 586 FDSLTPLEDEGPTVGTGTFELMS--SKDLAYQMTIYDWEIPLFCVHELELITFTGRRN 643
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 814 KEINSPRLKMHRTNLTLMFEKCIVENEFERAVLSRIEILQVQDINNENGYLE 873
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 644 FKTTANLDELFRFNEIQFWYVEICLSQSLSKRVQLKKFIKIAHCKEYKNLSFFA 703
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 874 IVSAVNSVSYRLDHTFEALQER-KRRILD-DA-VELSDHFEKKYLVKLKSNPCVPFF 931
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 704 IVMGLSNIAVSRLALTWELKPSKFKFYAFESLMPSRNH-RAYRLVATLEPPLIPFM 763
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 932 IYLTNLTKEGNSDFLKRKGDILNFSKRRVAETIGEIQYQONPY 979
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 764 PLLIKDMFTFHEGN-KTFI--DN-LVNFEKRMNANTARTVATYRSQPF 807
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT      9
ENTRY       S30356 #type complete

```



```

37-69      #domain SH3 homology #label SH3B\
1064-1305  #domain CD25-type guanine nucleotide exchange activator\
           homology #label SOS\
SUMMARY    #length 1333 #molecular-weight 152115 #checksum 7595

```

D6 1071 ELAROTTEFLFY - CKTKRACIAKWGKSGSLESIDSTIOPIKASNOJTNVGYM- 1128
 :|||:: |:: :: ||
 QY 613 DLAQCITIVDMELFNC-VHELELYHFEGHN-FKTTANDELFRRENEIQWVYTEIC 670

1100 1094 LAGATLADTALDQDNHNDNMTSLTTLTGSSGNLFFNNETENLNL 1100

671 LC-SQLSKRVQLLKFKFIKIAHCKEYKNLNSFPAIVMGSLNIAVSRLLATWEKLPSEKFK 729

```

      : : | : | | | | | | | | | |
730 FYAEFESLMDPSRNHRAVRLTVAK LEP-PLIFMPLLIKMTFTHEGKNTFI-D-N-LV 784
      : : | : | | | | | | | | | |
QY

```

```

785 NFEKRMIANTARTVRYRSQPN-PDAQANK 816
QY

```

RESULT	10	
ENTRY	S25716	#type complete
TITLE	Ras guanine nucleotide exchange factor son-of-sevenless (sos)	
	1 - mouse	

ALTERNATE_NAMES	probable ras activator
ORGANISM	#formal_name Mus musculus #common_name house mouse
DATE	13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change

ACCESSIONS	S25716; S21391
REFERENCE	S25714
#authors	Bowtell, D.; Fu, P.; Simon, M.; Senior, P.

#title	Identification of murine homologues of the Drosophila Son of sevenless gene: potential activators of ras.
#cross-references	MUID:92335328

```
##molecule_type mRNA
##residues 1-1336 ##label BOW
##cross-references EMBL:Z11574; NID:g54134; PTD:g54135
```

FEATURE	
459-561	#domain pleckstrin repeat homology #label PLK\homology; pleckstrin repeat homology

```

793-1036      #domain CDC25- type guanine nucleotide exchange activator
               homology #label SOS
SUMMARY      #length 1336 #molecular-weight 152762 #checksum 9720

```

```
Query Match: 4.1%; Score 25; Dn 2; Length 136;  
Best Local Similarity 25.4%; Pred. No. 9, DE=21;  
Matches 58; Conservative 64; Mismatches 98; Indels 8; Gaps 6;
```

D**b** 771 FQSSPTVEWHISRPCHETEDDLTLPHBETIARQLTLESIDYRAVOPSELYGSWTIKED 830
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :
O**y** 586 FDSLTLPPEDEGFPVTGVOTFELMS--SKDIAYOMTIYDMEFLNCHEHLELTHFRGRHN 643
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :
D**b** 831 KEINSPRLKMIHTNTNLIMFEKGIVTEYLEBRVAVSRIEILIOVBOELNNNGGLE 890
 : : | : | : | : | : | : | : | : | : | : | : | : | : | :
O**y** 644 FKKTANMDELRRRNEIQFWWYEITCLCSLSKRVDLLKKFKLIANKCKEKMNISFFA 703
 : : | : | : | : | : | : | : | : | : | : | : | : | : | :
D**b** 891 VFSAMNSPPYRLDHTFEQIPSROKILLIEAHELSE-DHYKK-VIAKLRSINPCCVPFEG 948
 | : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
O**y** 704 IVMGLSINAISRATLWEKPLKEFKFAAEESIMDSPRNHRNRYRLTVAKLEPPLIFMP 763
 | : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
D**b** 949 TYLNLIKLVTEGPNREYLARKHGCELINFEXRKRVAVETIGELOOVONOPY 996
 : : : : | : | | : | : | : | : | : | : | : | : | : | : | : | :
O**y** 764 LLDMDTFTEHGN-KTFT--DV-LNFVKMKMLANTARKYVRKSDF 807

RESULT	11
ENTRY	RGBYC5
TITLE	#type complete cell division control protein CDC5 - yeast (Saccharomyces

ALTERNATE_NAMES	protein L2142.6; protein YLR310c
ORGANISM	#formal_name <i>Saccharomyces cerevisiae</i>
DATE	31-Mar-1988 #sequence_revision 31-Mar-1993 #text_change

ACCESSIONS	
REFERENCE	
#authors	
A26596; S51442; A23444; S43051; S47990	
A26596	
Brock, D.; Toda, T.; Michaeil, T.; Levin, L.; Bitchmeier, C.; Zoller, M.; Powers, S.; Wisler, M.	

#title The S. cerevisiae CDC25 gene product regulates the RAS/adenylylate cyclase pathway.

```
##molecule_type DNA
##residues 1-1589 ##label BRO
```

```
#authors      Pauley, A.
#submission  submitted to the EMBL Data Library, November 1994
```

```
##molecule_type DNA
##presidues 1-1589 ##label PAU
##cross-references EMBL:U17247; NID:g577216; PID:g577222; MIPS:YL3106
REFERENCE
A23444
```

#authors Camonis, J.H.; Kalekine, M.; Gondre, B.; Garreau, H.;
Boy-Marcotte, E.; Jaquet, M.
#journal EMBO J. (1986) 5:375-380

gene which controls the cyclic AMP level of *Saccharomyces cerevisiae*.
#cross-references MUID:86220116

```
##molecule_type DNA
##residues 1-496, 'Y', 498-953, 'LSYIMNLSR', 964-1589 ##label CAM
##cross-references EMBL:X03579; NID:g3483; PID:g3484
```

#authors	Daniel, J.H.
#journal	Curr. Genet. (1986) 10:879-885
#title	The CDC25 "Start" gene of <i>Saccharomyces cerevisiae</i> :

homologies with rhodopsin and cytochrome P450.
#cross-references MUID:88194639
#accession S43051

```
##molecule_type DNA
##residues 877-1589 ##label DAN
##cross-references EMBL:X03579
```

GENETICS

gene	SGD:CDG25: CTN1
##cross-references	SGD:S0004301; MIPS:YLR310c
map_position	12R
FUNCTION	
description	positive control of level of cellular cAMP at the stage at which the cell division cycle is triggered
CLASSIFICATION	##superfamily budding yeast CDG25: CDG25-type guanine nucleotide exchange activator
KEYWORDS	cell cycle control; transmembrane protein
FEATURE	
65-123	domain SH3 homology #label SH3\
1301-1542	domain CDG25-type guanine nucleotide exchange activator homology #label SOS
SUMMARY	length 1589 #molecular-weight 179090 #checksum 7509
Query Match	4.1% Score 255; DB 1; Length 1589;
Best Local Similarity	25.5% Pred. No. 9,726-21;
Matches	71; Conservative 77; Mismatches 117; Indels 13; Gaps 10;
Db	1228 VSEKIPGAEDLQKINIKELINENEKPPVDPKQDSVSAVOTTKRDNKSPI-HMS5-SSL 1285
Qy	537 VADKLSGEBELI-IV-KMSSGEKKVVLKPNDV-SVETTLTNGRLFACPREPQDSTPL 592
Db	1286 PSSASSAFPLKLLKLLDDPTVTAQTLVLEHDIYLRITNPECCIDRMRGKYYCNMGSP 1345
Qy	593 PEQGPVGVGVGFELMSSKDLAY-QMTIYWEVENCHELELYHFGGHNKFK-TA 649
Db	1346 NITKFINANTLINFVSHTRKQADYKRSKLTQYFVTVACHCKELNFSMATAVSALY 1405
Qy	650 NLDLFLRFHEFQWVYTELCOSQSKRVQGLKFKIKIAACHCKEYKVLNFSFAVMGLS 709
Db	1406 SSPIYRLKTKWDLVSTESKDLKLNLMDSKRNKYKRELLRSTVDVACVPPFGVLSL 1465
Qy	710 NIAVSRLATLWEKLPSEKFKFYAEFESLMDPSNHRAR-LTVAKLEPLLIPEMPLIKD 768
Db	1466 LTFEFGNPDLHNSTNINFSKRTIANIVEIISIF 1503
Qy	769 MTFHEGKRTFI-D-NLVNFEKMRMIAMTARYRYR 803
RESULT	12
ENTRY	A41216
TITLE	guanine nucleotide exchange activator son-of-sevenless - fruit fly (Drosophila melanogaster)
ORGANISM	#formal_name Drosophila melanogaster
DATE	28-May-1992 #sequence_revision 28-May-1992 #text_change 24-Sep-1998
ACCESSIONS	A41216
REFERENCE	A41216
authors	Simon, M.A.; Bowtell, D.D.L.; Dodson, G.S.; Lavery, T.R.; Rubin, G.M.
#journal	Cell (1991) 67:701-716
#title	Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase
#cross-references	MUID:92034991
#accession	A41216
##status	preliminary
##molecule_type	mRNA
##residues	1-1596 ##label SIM
##cross-references	GB:M7501; NID:G158470; PID:G158471
GENETICS	FLYBase:Sos
CLASSIFICATION	##cross-references FLYBase:PB9n0001965
homology	pleckstrin repeat homology
FEATURE	
479-586	domain pleckstrin repeat homology #label PLK\
825-1066	domain CDG25-type guanine nucleotide exchange activator homology #label SOS
SUMMARY	length 1596 #molecular-weight 177933 #checksum 4040
Query Match	4.0% Score 251; DB 2; Length 1596;

	Best local Similarity	24.4%	Pred.No.	4.22e-20;	
Matches	48;	Conservative	59;	Mismatches	86;
				Indels	4;
				Gaps	3;
Db	832	ELARQTLLEFEKKNKVKSELYGSPWTKKQKVKSPULKIMHNTVNVRIEKSITEA	891		
Qy	613	DLAYQMTYDWELEFNCVHLEELIYTPGGRNPKTTALDLFLRNFMEIQFWYTELCLC	672		
Db	892	ENYEERLAIQRAIEVMYVMEELNNFNGILSLVAAMGTAAYRLRLRWFGQGLPEYKRFLE	951		
Qy	673	SQSKRQVLKTKIKIAACHCKEYKKNLSEFALVGLSLMIAVSRLATLWEEKLPKSPKFFYA	732		
Db	952	ECRELSD-DHLK-YOERLRSLNPPCVPEFGFGRYLLNIIHLGEGNDLLANTELINESKR	1009		
Qy	733	EPESIMDPSSNNHAYRLVAKLEPLIPMLIKDMFTFEGNKKTFIDN-LVNEFKMR	790		
Db	1010	KVAEIIIGIDQYONOPY	1026		
Qy	791	MIANTARTVRYYSQPF	807		

```

ENTRY      13
TITLE      S64758 #type complete
ALTERNATE_NAMES
ORGANISM   SCD25 protein (version 2) - yeast (Saccharomyces cerevisiae)
DATE       protein L1309; protein YIL016w
          #formal_name Saccharomyces cerevisiae
          01-May-1995 #sequence_revision 24-May-1996 #text-change
          06-Feb-1998
ACCESSIONS
REFERENCE  S64758; S64764; S69390; S70559
#authors   Miosga, T.; Zimmermann, F.K.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64758
##molecule_type DNA
#residues  1-1048 ##label MTO
#cross-references EMBL:Z73121; NID:g1360186; PID:e245452; PID:g1360187;
          MIPS:YIL016w
REFERENCE  #experimental_source strain S288C
#authors   Goffeau, A.; Purnelle, B.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64764
##molecule_type DNA
#residues  1-1048 ##label GOF
#cross-references EMBL:Z73121; NID:g1360186; PID:e245452; PID:g1360187;
          MIPS:YIL016w
##experimental_source strain S288C
REFERENCE  S69380
#authors   Purnelle, B.; Goffeau, A.
#submission submitted to the EMBL Data Library, April 1996
#description The sequence of 32 kb on the left arm of yeast chromosome XII
          reveals 14 open reading frames among which HSP104, SSA2,
          SPA2, KNS1, DPS1/APS, SDC25, a new member of the
          serpinuperin family and a new ABC transporter homologous
          to the human multidrug resistance protein.
#accession S69390
##molecule_type DNA
#residues  1-1048 ##label PUR
#cross-references EMBL:X97560; NID:g1297003; PID:e238680; PID:g1297014
          S70557
REFERENCE  S70557
#authors   Miosga, T.; Zimmermann, F.K.
#journal    Yeast (1996) 12:693-708
#title      Sequence analysis of the CEN12 region of Saccharomyces
          cerevisiae on a 4.7 kb fragment of chromosome XII
          including an open reading frame homologous to the human
          cystic fibrosis transmembrane conductance regulator protein
          CTRR.
#cross-references MUD:96405918
#accession S70559
#status     nucleic acid sequence not shown; translation not shown
##molecule_type DNA
#residues  1-1048 ##label MIV
#cross-references EMBL:X91488; NID:g1495203; PID:e199006; PID:g1495207
#note       The nucleotide sequence was submitted to the EMBL Data

```

CLASSIFICATION	#superfamily CDC25-type guanine nucleotide exchange activator
FEATURE	homology
946-1197	#domain CDC25-type guanine nucleotide exchange activator homology #label SOS
SUMMARY	#length 1250 #molecular-weight 144779 #checksum 1783
Query Match	3.8%; Score 236; DB 2; Length 1250;
Best Local Similarity	33.0%; Pred. No. 9,82e-18;
Matches	35; Conservative 26; Mismatches 44; Indels 1; Gaps 1;
Db	1018 SKRAKILSHFIAEVCRCRKNFSSMTDIIAALSSPYRLKTKQAVIPQTRDLQSLN 1077
OY	676 SKRVQLKKFKTIKAHCKEYKNNISFFAIVMLISIAVSALTLMTKLPKSKFKFYAEPE 735
Db	1078 KLMDPKKFNINRNDELKSLASACVPEFCGYSDLTFDSCGNPYL 1123
OY	736 SLMDPSRNRARLVAKLP-PLIFPMFLLIKDMTFTHBGNKTFI 780
RESULT	15
ENTRY	ORGANISM
TITLE	protein kinase (EC 2.7.1.37), CAMP-dependent, type I regulatory chain - California sea hare
ORGANISM	#formal_name Aplysia californica #common_name California sea hare
DATE	31-Mar-1993 #sequence-revision 31-Mar-1993 #text-change 03-Sep-1997
ACCESSIONS	JH0590
REFERENCE	Bergold, P.J.; Beushausen, S.A.; Sacktor, T.C.; Cheley, S.; Bayley, H.; Schwartz, J.H. Neuron (1992) 8:387-397
#authors	A regulatory subunit of the CAMP-dependent protein kinase down-regulated in aplysia sensory neurons during long-term sensitization.
#journal	
#title	
#cross-references	MUID:92135429
#accession	JH0590
#molecule_type	mRNA
##residues	1-378 ##label BER
##cross-references	EMBL:X62382; NID:95587; PID:95588
##note	this protein is identified as the M4 isoform
COMMENT	The inactive form of the enzyme is composed of two regulatory chains and two catalytic chains. Activation by CAMP produces two active catalytic monomers and a regulatory dimer that binds four CAMP molecules.
COMMENT	Type I regulatory chains contain a high-affinity binding site for MgATP.
COMMENT	The pseudophosphorylation site binds to the substrate-binding region of the catalytic chain but is not phosphorylated. The physiological significance of phosphorylations by other kinases is unclear.
CLASSIFICATION	#superfamily CAMP-dependent protein kinase regulatory chain; homology
KEYWORDS	acetylated amino end; ATP binding; CAMP binding; duplication; heterotetramer; homodimer; phosphoprotein; phosphotransferase
FEATURE	
2-378	#product protein kinase, CAMP-dependent, type I-beta regulatory chain #status predicted #label M4V
2-133	#domain protein interaction #label DMV
93-97	#region pseudophosphorylation motif
134-251	#domain CAMP receptor protein cyclic nucleotide-binding domain homology #label CA1
252-378	#domain CAMP receptor protein cyclic nucleotide-binding domain homology #label CA2
2	#modified site acetylated amino end (Ala) (in mature form) #status predicted
199,208	#binding_site CAMP (Glu, Arg) #status predicted
323,332	#binding_site CAMP (Glu, Arg) #status predicted
SUMMARY	#length 378 #molecular-weight 42737 #checksum 2977

Sun Apr 23 09:32:16 2000

US-09-422-999-18.rpr

Page 8

Query Match	3.58;	Score 216;	DB 1;	length 378;
Best Local Similarity	32.7%;	Pred. No. 1.21e-14;		
Matches 33;	Conservative 26;	Mismatches 39;	Indels 3;	Gaps 3;
Db	135	FSHLDNERSDIDAM-FVPHAGEVILIOGDGDMFYVIOGGEYDVV-NNVVTSIG	192	
		::: :::		
Qy	195	LSHSTIVKELAGVILFESHAKGSTVLFENOGEGISWTIIKGSNNVYIYKGVYCTIH	254	
Db	193	EGGSFEGELIVGTGPRATVKAKTQ-KLMDIDSDSYRIL	232	
		::: :		
Qy	255	EGDDEGRLALVNDAPRAASIVLRDNCHEFLVQKDEPNRIL	295	

Search completed: Sat Apr 22 13:19:19 2000
Job time : 39 secs.

MWSEKELI
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat Apr 22 13:23:13 2000; MasPar time 14.86 Seconds
Tabular output not generated. 759.883 Million cell updates/sec

Title: >US-09-422-999-18
Description: (1-849) from US09422999.pep
Perfect Score: 6222
Sequence: 1 MVLKRALVTGFLTRTHL.....VIDNORTLSMSHRLPRRP 849

Scoring table: PAM 150
Gap 11

Searched: 134018 seqs, 13297625 residues

Post-Processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 36.052; Variance 179.682; scale 0.201

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	287	4.6	489	1	US-08-318-	Sequence 4, Applicatio	1.23e-12
2	287	4.6	666	1	US-08-318-	Sequence 3, Applicatio	1.23e-12
3	287	4.6	814	1	US-08-318-	Sequence 2, Applicatio	1.23e-12
4	276	4.4	402	2	US-08-290-	Sequence 14, Applicati	8.79e-12
5	276	4.4	652	1	US-08-318-	Sequence 8, Applicatio	8.79e-12
6	266	4.3	423	2	US-08-290-	Sequence 11, Applicati	5.22e-11
7	266	4.3	1287	2	US-08-290-	Sequence 4, Applicatio	5.22e-11
8	256	4.1	364	1	US-08-318-	Sequence 6, Applicatio	3.07e-10
9	255	4.1	423	2	US-08-290-	Sequence 10, Applicati	3.07e-10
10	255	4.1	426	2	US-08-290-	Sequence 12, Applicati	3.07e-10
11	255	4.1	1319	2	US-08-290-	Sequence 2, Applicatio	3.67e-10
12	255	4.1	1336	2	US-08-290-	Sequence 6, Applicatio	3.67e-10
13	251	4.0	430	2	US-08-290-	Sequence 9, Applicatio	7.43e-10
14	251	4.0	1572	2	US-08-290-	Sequence 5, Applicatio	7.43e-10
15	236	3.8	418	2	US-08-290-	Sequence 13, Applicati	1.03e-08
16	191	3.1	362	2	US-08-290-	Sequence 15, Applicati	2.39e-05
17	143	2.3	852	4	PCT-US95-0	Sequence 5, Applicatio	6.37e-02
18	143	2.3	852	4	PCT-US95-0	Sequence 5, Applicatio	6.37e-02
19	133	2.1	768	4	PCT-US95-0	Sequence 2, Applicatio	3.07e-01
20	133	2.1	768	4	US-08-408-	Sequence 2, Applicatio	3.07e-01
21	122	2.0	528	1	US-08-363-	Sequence 14, Applicati	1.67e+00
22	122	2.0	530	1	US-08-363-	Sequence 3, Applicatio	1.67e+00
23	122	2.0	530	1	US-08-363-	Sequence 10, Applicati	1.67e+00

24	122	2.0	530	1	US-08-363-	Sequence 9, Applicatio	1.67e+00
25	122	2.0	530	1	US-08-363-	Sequence 8, Applicatio	1.67e+00
26	122	2.0	530	1	US-08-363-	Sequence 2, Applicatio	1.67e+00
27	120	1.9	210	2	US-08-892-	Sequence 2, Applicatio	2.26e+00
28	116	1.9	451	1	US-08-625-	Sequence 4, Applicatio	4.13e+00
29	119	1.9	1159	2	US-08-956-	Sequence 13, Applicati	2.63e+00
30	106	1.6	888	2	US-08-956-	Sequence 4, Applicatio	1.81e+01
31	98	1.6	178	1	US-07-929-	Sequence 2, Applicatio	5.68e+01
32	98	1.6	184	2	US-08-945-	Sequence 10, Applicati	5.68e+01
33	97	1.6	442	1	US-08-363-	Sequence 11, Applicati	6.53e+01
34	97	1.6	445	1	US-08-363-	Sequence 5, Applicatio	6.53e+01
35	99	1.6	480	1	US-07-803-	Sequence 2, Applicatio	4.93e+01
36	101	1.6	543	2	US-08-922-	Sequence 10, Applicati	3.71e+01
37	97	1.6	858	1	US-07-712-	Sequence 2, Applicatio	6.53e+01
38	102	1.6	1167	1	US-08-485-	Sequence 6, Applicatio	3.21e+01
39	102	1.6	1167	2	US-09-184-	Sequence 6, Applicatio	3.21e+01
40	102	1.6	1167	2	US-08-590-	Sequence 6, Applicatio	3.21e+01
41	102	1.6	1168	1	US-08-620-	Sequence 9, Applicatio	3.21e+01
42	98	1.6	2319	1	US-08-212-	Sequence 8, Applicatio	5.68e+01
43	98	1.6	2319	4	PCT-US94-1	Sequence 6, Applicatio	5.68e+01
44	98	1.6	2319	1	US-08-474-	Sequence 6, Applicatio	5.68e+01
45	98	1.6	2319	2	US-08-670-	Sequence 6, Applicatio	5.68e+01

ALIGNMENTS

RESULT	1	STANDARD	PRT	489 AA.
ID	US-08-318-831-4			
XX	xxxxxx			
AC				
XX				
DT				
XX				
DE	Sequence 4, Application US/08318831			
XX				

Sequence 4, Application US/08318831
Patent No. 5656595
GENERAL INFORMATION:
APPLICANT: Schweighofer, Fabien
TITLE OF INVENTION: PEPTIDES HAVING A GDP EXCHANGE FACTOR
TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SAID PEPTID
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Road, 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: System 7.1
SOFTWARE: Word 5.1 (EPO Patentin)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318, 831
FILING DATE: 19 October 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR92/04827
FILING DATE: 21-Apr-1992
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Julie K.
REGISTRATION NUMBER: P-38, 619
REFERENCE/DOCKET NUMBER: ST92033-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 454-3839
TELEFAX: (610) 454-3808
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids

[illegible]

QY 656 RRENIOFWVTEICLSQSLSKRVQLKKFIKIAHCKEYKNNSPFAYMGISNTAVSR 715
DB 685 LKKTWIKSKOKRALIDKIQIVSSEGRFKNLRBALNCDDPCVPYIGMTLDLAFIEEG 744
QY 716 LALTWEKLPSEKKEKYAFESIMPSRNRHAYRLTVAKLEPPLIPFMDLIKMDITFHBG 775
DB 745 TPVYTEDGLVNFSGKRMISHIREIROPQOYAKIE 780
QY 776 NKTFTDN-LVNFEMKRMIAHTARYRYRSQFPNPD 810

RESULT 4
ID US-08-290-731C-14 STANDARD; PRT; 402 AA.
XX xxxxxx
XX

DE Sequence 14, Application US/08290731C
XX
CC Sequence 14, Application US/08290731C
CC Patent No. 5843646
CC GENERAL INFORMATION:
CC APPLICANT: BOWTELL, David Douglas Lawrence
CC TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
CC TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
CC TITLE OF INVENTION: AND MSOS POLYPEPTIDES
CC NUMBER OF SEQUENCES: 15
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
CC STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CC CITY: WASHINGTON
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20037
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,731C
CC FILING DATE: 17-OCT-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/AU93/00068
CC FILING DATE: 17-FEB-1993
CC PRIOR APPLICATION DATA: PLO921/92
CC APPLICATION NUMBER: PLO921/92
CC FILING DATE: 17-FEB-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KIT, Gordon
CC REGISTRATION NUMBER: 30,764
CC REFERENCE/DOCKET NUMBER: O-36066
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 293-7060
CC TELEFAX: (202) 293-7860
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 402 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 402 AA; 844812 CN;

Query Match 4.4%; Score 276; DB 2; Length 402;
Best Local Similarity 22.2%; Pred. No. 8,79e-12;
Matches 57; Conservative 75; Mismatches 119; Indels 6; Gaps 6;

DB 146 QLEEDLS-LHNSPDPPIYDELVLPPREIAKQCLIEPFSFHSIRIOPFIRKINDELN 204
QY 584 EDPDSILPLPEORGPVGTGTFELMSSKDLAYQWTIYDWELFNCVHELILYHTFGRHN 643

DB 205 RFSPEKSTFELYLN-HLVNF-VTEIVQEEPRRRINVLAFYIQVDYLBELNPSASF 262
QY 644 -FKRTANLDLEFLRRNEIOFWVTEICLSQSLSKRVQLKKFIKIAHCKEYKNNSEF 702
DB 263 SITALSNSPPIHLEKRTWMLNLSKTLASPELLNTEARKNFSVYRCLENCVPCVPL 322
QY 703 AYWGLSNINAVSLALTWEKLPSEKKEKYAFESIMDSRNRHAYRLTVAKLEPPLIFEM 762
DB 323 GYVFTDLTFELTKGNKDNFQNMINDKRTKVRILNEIKFQSGVGMFNPINEVOELINEV 382
QY 763 PLIKMDTFEHSNKTFIDNLVNFEMKRMIAHTARYRYRSQFPNDAQANKNH-QDV 821
DB 383 ISREERNNTI-YORSLT 398
QY 822 RSYROLNVTIDNORTIS 838

RESULT 5
ID US-08-318-831-8 STANDARD; PRT; 652 AA.
XX xxxxxx
XX

DE Sequence 8, Application US/08318831
XX
CC Sequence 8, Application US/08318831
CC Patent No. 5656595
CC GENERAL INFORMATION:
CC APPLICANT: Schweigboffer, Fabien
CC TITLE OF INVENTION: Peptides Having a GDP Exchange Factor
CC TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SAID PEPTID
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Rhone-Poulenc Rorer Inc.
CC STREET: 500 Arcoia Road, 3043
CC CITY: Collegeville
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19426
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: Macintosh
CC OPERATING SYSTEM: System 7.1
CC SOFTWARE: Word 5.1 (Epo PatentIn)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/318,831
CC FILING DATE: 19 October 1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: FR97/04827
CC FILING DATE: 21-APR-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Julie K.
CC REGISTRATION NUMBER: P-38,619
CC REFERENCE/DOCKET NUMBER: ST92033-US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (610) 454-3839
CC TELEFAX: (610) 454-3808
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 652 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC SEQUENCE 652 AA; 74235 MW; 2320730 CN;

Query Match 4.4%; Score 276; DB 1; Length 652;
Best Local Similarity 25.8%; Pred. No. 8,79e-12;
Matches 59; Conservative 71; Mismatches 89; Indels 10; Gaps 7;

RESULT	6	
ID	US-08-290-731C-11	STANDARD; PRT; 423 AA
yy		

Sequence 11, Application US/08290731C

Sequence 11, Application US/08290731C
Patent No. 5843646

APPLICANT: BOWTELL, David Douglas Lawrence
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
TITLE OF INVENTION: SON OF SEVENTEENLESS (MSOS) GENE,
TITLE OF INVENTION: AND MSOS POLYPEPTIDES
NUMBER OF SEQUENCES: 15

ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.

STATE: D.C.
COUNTRY: USA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00068

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PL0921/92

ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon

REFERENCE/DOCKET NUMBER: Q-36066
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 233-7000
TELEFAX: (202) 293-7860
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 423 amino acids

LIFE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

Query Match 4.38; Score 266; DB 2; Length 423;

Best Local Similarity 24.9%; Pred. No.5,22e-11;
Matches 57; Conservative 70; Mismatches 92; Indels 10; Gaps 7.

RESULT	7		
ID	US-08-290-731C-4	STANDARD;	PRT; 1297 AA

Sequence 4, Application US/08290731C

Sequence 4, Application US/08290731C
Patent No. 5843646

APPLICANT:	BOWTELL,	David Douglas Lawrence
TITLE OF INVENTION:	DNA MOLECULES ENCODING	MURINE
TITLE OF INVENTION:	SON OF SEVENTEENLESS (MSOS)	GENE.
TITLE OF INVENTION:	AND MSOS POLYPEPTIDES	
NUMBER OF SEQUENCES:	15	

ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.

STATE: D.C.
COUNTRY: USA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU93/00068

FILING DATE: 17 FEB 1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PL0921/92

FILED DATE: 17 FEB 1992
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon

REGISTRATION NUMBER: 30/704
REFERENCE/DOCKET NUMBER: Q-36066
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 253-7000
TELEFAX: (202) 293-7860
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 4
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1297 amino acids

MOLECULE TYPE: protein

Seq	SEQUENCE	1297 AA; 148846 MW; 9059290 CN;
Query Match		4.3%; Score 266; DB 2; Length 1297;
Best Local Similarity		24.9%; Pred. No. 5.22e-11;
Matches	57; Conservative	70; Mismatches 92; Indels 10; Gaps 7;
Db	717 FESSPPEVBMHISPRGQETPDMLPHIPETAROLTLLESDLYRKVQSGSELVGSWTKEED 776	
Oy	566 FDSLTPPEQSGPYVGTGTEFLMS--SKDLAYQMTIYDWELENFNCVHELELLYHTFGRRN 643	
Db	777 KEINSPNLKMKIRHTNTLTLMEFKCIVEAEENFEENVAVLSRIVEILQYFQDILNNGYLE 836	
Oy	644 FKKTANIDLFLRRNNEIQFWYVEIICLSQSLSKRVOLKFKIKIAAHCKEKNINSFPA 703	
Db	837 IVSAVSVSVYRDLHTEFALQER--KRRIID-DA-VELSQDHEKKYLVKLSINPCEVPEF 893	
Oy	704 IVMGSLNIAVSRALATWKEPLPSFKFKFAEFESLMDPSRNH--RAYRLVIAKLEPLPIPFM 762	
Db	894 GIYLTNIIKTEGNSDPLKRRCKDLINFSKRKVAEITGEIQYONQY 942	
Oy	763 PLLIKDMTFTEGKNTFT--D-NLVNFEKMKMIANTARTVYKRSOPF 807	
RESULT	8	
ID	US-08-318-831-6	STANDARD; PRT; 364 AA.
AC	xxxxxx	
DT		
XX		
XX		
Sequence 6, Application US/08318831		
CC	Sequence 6, Application US/08318831	
CC	Patent No. 5656595	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Schweighoffer, Fabien	
CC	APPLICANT: Tocque, Bruno	
CC	TITLE OF INVENTION: PEPTIDES HAVING A GDP EXCHANGE FACTOR	
CC	TITLE OF INVENTION: ACTIVITY, NOCLEIC ACID SEQUENCES CODING FOR SAID PEPTIDES	
CC	TITLE OF INVENTION: PREPARATION AND UTILIZATION	
CC	NUMBER OF SEQUENCES: 12	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Rhone-Poulenc Rorer Inc.	
CC	STREET: 500 Arcola Road, 3C43	
CC	CITY: Collegeville	
CC	STATE: PA	
CC	COUNTRY: USA	
CC	ZIP: 19426	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Floppy disk	
CC	COMPUTER: Macintosh	
CC	OPERATING SYSTEM: System 7.1	
CC	SOFTWARE: Word 5.1 (Epo Patentin)	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: US/08/318,831	
CC	FILING DATE: 19 October 1994	
CC	CLASSIFICATION: 435	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: FR92/04827	
CC	FILING DATE: 21-APR-1992	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Smith, Julie K.	
CC	REGISTRATION NUMBER: P-38,619	
CC	REFERENCE/DOCKET NUMBER: ST92033-US	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (610) 454-3839	
CC	TELEFAX: (610) 454-3808	
CC	INFORMATION FOR SEQ ID NO: 6:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 364 amino acids	
CC	TYPE: amino acid	
CC	TOPOLOGY: linear	
SEQ	SEQUENCE 364 AA; 41968 MW; 717826 CN;	

	Query Match	4.1%	Score 256;	DB 1;	length 364;
Bc	Best Local Similarity	25.4%	Pred. No. 3.07e-10;		
Matches	58;	Conservative	65;	Mismatches 97;	Indels 8; Gaps 6;
Db	22 FQSPPEVEMWISPPGHIEFPDLTLPHIEIAROLTLDSIDLKRAVOPSDVGSVWTKE	81			
Oy	586 FDSLPLPDEGCPGVGVTFEELMS--SKDLAYGMTIYDWELFNCFVHELIYHTGRHN	643			
Oy	82 KEINSPILKAIRHTNTLITLMFECCYEYETENLERAVAVSRILEILOVFOELNNGVLE	141			
Oy	644 FKKTANMIDFLRRFNEJOFVWVEICLCISOLSKRYOLLKKFIKAHCKEYNLSFFA	703			
Db	142 VVSANSSPVYRLDTHTEPOIRSROKLTLEAHKSE--DHKK-VLAATRSINPCVFFG	199			
Oy	704 IVMGISIAVSRLLATWEKLPSKKRKYAEESIMDSRNHRATRLTVAKDELPIIFMP	763			
Db	200 IYLNIHLTEGNPEVLKRHGKELINFSKRRKVAETGEILOOYONCPY	247			
Oy	764 LLIKMFETHEGN-KTFL--DN-LVNEKMMMIANTARTVRYNSQPF	807			
RESULT	9				
ID	US-08-290-731C-10	STANDARD;	PRT;	423 AA.	
AC	xxxxxx				
Dt					
De	Sequence 10, Application US/08290731C				
XX					
XX					
CC	Sequence 10, Application US/08290731C				
CC	Patent No. 5843646				
CC	GENERAL INFORMATION:				
CC	APPLICANT: BOWTEL, David Douglas Lawrence				
CC	TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE				
CC	TITLE OF INVENTION: SON OF SEVENTELESS (MSOS) GENE,				
CC	TITLE OF INVENTION: AND MSOS POLYPEPTIDES				
CC	NUMBER OF SEQUENCES: 15				
CC	CORRESPONDENCE ADDRESS:				
CC	ADDRESSEE: SUGHRUE, MTOM, ZINN, MACPEAK & SEAS				
CC	STREET: 2100 PENNSYLVANIA AVENUE, N.W.				
CC	CITY: WASHINGTON				
CC	STATE: D.C.				
CC	COUNTRY: USA				
CC	ZIP: 20037				
CC	COMPUTER READABLE FORM:				
CC	MEDIUM TYPE: Floppy disk				
CC	COMPUTER: IBM PC compatible				
CC	OPERATING SYSTEM: PC-DOS/MS-DOS				
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25				
CC	CURRENT APPLICATION DATA:				
CC	APPLICATION NUMBER: US/08/290,731C				
CC	FILING DATE: 17-OCT-1994				
CC	CLASSIFICATION: 435				
CC	PRIOR APPLICATION DATA:				
CC	APPLICATION NUMBER: PCT/AU93/00068				
CC	FILING DATE: 17-FEB-1993				
CC	PRIOR APPLICATION DATA:				
CC	APPLICATION NUMBER: PL0921/92				
CC	FILING DATE: 17-FEB-1992				
CC	ATTORNEY/AGENT INFORMATION:				
CC	NAME: KIT, Gordon				
CC	REGISTRATION NUMBER: 30,764				
CC	REFERENCE/DOCKET NUMBER: Q-36066				
CC	TELECOMMUNICATION INFORMATION:				
CC	TELEPHONE: (202) 293-7060				
CC	TELEFAX: (202) 293-7860				
CC	TELEX: 6491103				
CC	SEQUENCE CHARACTERISTICS:				
CC	SEQUENCE FOR SEQ ID NO: 1:				
CC	LENGTH: 423 amino acids				
CC	TYPE: amino acid				


```
CC      FILING DATE: 17-FEB-1992
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: KIT, Gordon
CC      REGISTRATION NUMBER: 30,764
CC      REFERENCE/DOCKET NUMBER: Q-36066
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (202) 293-7060
CC      TELEFAX: (202) 293-7860
CC      TEXT: 6491103
CC      INFORMATION FOR SEQ ID NO: 6:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 1336 amino acids
CC      TYPE: amino acids
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 1336 AA; 152734 MW; 9610574 CN;
SQ
Query Match          4.1%; Score 255; DB 2; Length 1336;
Match Local Similarity 25.4%; Pred. No. 3.67e-10;
Matches    58; Conservative    64; Mismatches 98; Indels   8; Gaps   6;

Db     771 FQSSPVEWEMISRGHIEFDLTLHPHEIAROLFTLESPLRYAVOPSELVSWTKED 830
OY     566 FDSLRLPEEGSPGVGYGTFFELMS--SKDLAYQMITYDWLEFNCVHELLEYTFEGRHN 643
Db     831 KEINSPILKMRHTNTLTLMFEKCIYETENLEERVAVSRKIIEILOVFDELNNFGYLE 890
OY     644 EFKTIANDELFRNEHQFWVTVEICISQISRVOYLKKFIKIAHCKEYKNLSNFA 703
Db     891 VYSANNSPPVRLDTHFPQIRPSROKLIIEAHELSE-DHYKK-VLAIKRSINPCVPFFG 948
OY     704 IVMGISNAIVSRLATLVWKLPSSKRKFYAEEESLMDDSRNRHRAVRLTVAKLEPLLFPMP 763
Db     949 IYLTNLKTGEANGPEVLRRHGKELINEFSKRRVAEITGEIQOYONOPY 996
OY     764 LLIKMTTFHSGN-KTFI--DN-LVNFEKMKMIANTARTVAYVRSQPF 807

RESULT 13
XX      US-08-290-731C-9 STANDARD; PRT; 430 AA.
AC      xxxxxx
DX
DY
XX
DE
Sequence 9, Application US/08290731C
XX      Sequence 9, Application US/08290731C
CC      Patent No. 5843646
CC      GENERAL INFORMATION:
CC      APPLICANT: BOWTELL, David Douglas Lawrence
CC      TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
CC      TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
CC      TITLE OF INVENTION: AND MSOS POLYPEPTIDES
CC      NUMBER OF SEQUENCES: 15
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS
CC      STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CC      CITY: WASHINGTON
CC      STATE: D.C.
CC      COUNTRY: USA
CC      ZIP: 20037
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/290,731C
CC      FILING DATE: 17-OCT-1994
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/AU93/00068
```

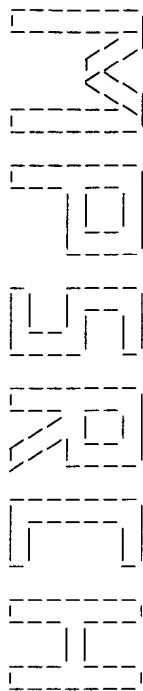

CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,731C
CC FILING DATE: 17-OCT-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/AU93/00068
CC FILING DATE: 17-FEB-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PLO921/92
CC FILING DATE: 17-FEB-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KIT, Gordon
CC REGISTRATION NUMBER: 30,764
CC REFERENCE/DOCKET NUMBER: Q-36066
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 293-7060
CC TELEFAX: (202) 293-7860
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 418 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 418 AA; 48715 MW; 949499 CN;

Query Match 3.8%; Score 236; DB 2; Length 418;
Best Local Similarity 33.0%; Pred. No. 1.03e-08;
Matches 35; Conservative 26; Mismatches 44; Indels 1; Gaps 1;

Db 235 SKRAKLSHPTFIAYCYGRKFNFSMTDIIISALYSPPIYRLEKTKWQAVIPQTRDLQSLN 294
Qy 676 SKRVOLKKFTKIAHCKEYKRLNSFEAIYWGLSNIAVSRLATLWTKLPKSKFKTYAEFE 735

Db 295 KIMDPKKNFINRNEIKSLHSAPCVPEFGVYLSDLFTDSGNPDYL 340
Qy 736 SIMDPSRNHRAVRLTVAKLEP-PLIPFWPLLIKDMFTFHGNGKTFPI 780

Search completed: Sat Apr 22 13:23:31 2000
Job time : 18 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 22 13:17:33 2000; Maspar time 44.93 Seconds

Tabular output not generated. 447.554 Million cell updates/sec

Title: >US-09-422-999-18
Description: (1-849) from US09422999.dep
Perfect Score: 6222
Sequence: 1 MVLKRALMTGFLTRRTL.....VINQRTLSQMSHRLPRRP 849

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 38.706; Variance 183.913; scale 0.210

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	370	5.9	1077	1	R91597	C3G protein.	1.53e-19
2	287	4.6	489	1	R50991	Peptide which modulate	9.88e-13
3	287	4.6	666	1	R50990	Peptide which modulate	9.88e-13
4	287	4.6	814	1	R43578	Peptide which modulate	9.88e-13
5	276	4.4	652	1	R43580	Peptide which modulate	7.60e-12
6	266	4.3	1297	1	R47048	Mammalian son of seven	4.81e-11
7	262	4.2	1297	1	R84539	MSO2 protein.	1.00e-10
8	256	4.1	364	1	R45379	Peptide which modulate	3.02e-10
9	255	4.1	1319	1	R47043	Mammalian son of seven	3.63e-10
10	255	4.1	1336	1	R84638	MSO2 protein.	3.63e-10
11	201	3.2	380	1	W54350	CAMP-dependent protein	5.94e-06
12	180	2.9	609	1	W87995	An alternatively splic	2.29e-04
13	180	2.9	728	1	W87994	A human MCG protein.	2.29e-04
14	134	2.2	932	1	W85040	PKG-green fluorescent	4.81e-01
15	134	2.2	933	1	W85039	Green fluorescent prot	4.81e-01
16	133	2.1	768	1	W00811	Ras p21 interacting pr	5.64e-01
17	122	2.0	528	1	W62086	Mycobacterium bovis	3.17e+00
18	122	2.0	528	1	W05831	M. tuberculosis RNA po	3.17e+00
19	122	2.0	530	1	R76480	Virulence-associated s	3.17e+00
20	122	2.0	530	1	W62085	Mycobacterium bovis	3.17e+00
21	116	1.9	451	1	W77095	Sorting nexin 2.	7.95e+00
22	117	1.9	485	1	Y00895	Rat GMEB-2' protein se	6.83e+00
23	117	1.9	529	1	Y00894	Rat GMEB-2' protein seq	6.83e+00

24	112	1.8	192	1	W55252	H. pylori ORF 05ep1171	1.45e+01
25	103	1.7	342	1	W55710	H. pylori ORF 06ep1061	5.49e+01
26	104	1.7	371	1	W01113	RNA polymerase primary	4.75e+01
27	104	1.7	382	1	W23969	Connexin protein Cx43.	4.75e+01
28	98	1.6	178	1	R26181	TSG-14.	1.13e+02
29	97	1.6	289	1	W43421	Rat syntaxin 3 protein	1.13e+02
30	102	1.6	449	1	W79140	Plant acetolactate syn	6.35e+01
31	99	1.6	480	1	R77249	Babesia microti p58.	9.77e+01
32	99	1.6	736	1	W97809	Human GTPase regulator	1.13e+02
33	97	1.6	858	1	R30477	Human leukocyteGF.	1.30e+02
34	102	1.6	1167	1	W31504	Nematode toxin 167p pr	6.35e+01
35	102	1.6	1167	1	W10653	Bacillus thuringiensis	6.35e+01
36	102	1.6	1168	1	W16326	Nematocidal toxin 167p	6.35e+01
37	98	1.6	2319	1	W53485	Mus factor VIII.	1.13e+02
38	98	1.6	2319	1	W44135	Mus musculus factor VI	1.13e+02
39	100	1.6	2555	1	W81512	Drosophila fat facets	8.47e+01
40	96	1.5	590	1	W94458	Human neurofibromatosis	1.49e+02
41	96	1.5	596	1	W94459	Human neurofibromatosis	1.49e+02
42	96	1.5	740	1	R68838	Plasmodium falciparum	1.49e+02
43	96	1.5	798	1	R57283	Bovine enterokinase.	1.49e+02
44	96	1.5	1086	1	W81839	Human LFA-1 alpha chain	1.49e+02
45	96	1.5	3457	1	W84560	Polypeptide encoded by	1.49e+02

ALIGNMENTS

RESULT	1	
ID	R91597	standard; Protein; 1077 AA.
AC	R91597;	
DT	26-JUL-1996	(first entry)
DE	C3G protein.	
KW	C3G; ras protein guanine nucleotide exchange factor; diagnosis; treatment; malignant tumour; activation.	
KW	treatment; malignant tumour; activation.	
OS	Homo sapiens.	
PN	U07051073-A.	
PD	28-FEB-1995.	
PF	13-JUN-1994; 130699.	
PR	11-JUN-1993; JP-140806.	
PA	(SHK7) SHINGIJUMSU JICODAN.	
DR	WPI; 95-127357/17.	
DR	N-PSDB; T13347.	
PT	Ras protein guanine nucleotide exchange factor C3G gene - useful for diagnosis and treatment of malignant tumours associated with ras oncogene activation.	
PS	Claim 9; Page 7-8; 9pp; Japanese.	
CC	The present sequence is that of C3G protein which is a ras protein	
CC	guanine nucleotide exchange factor. The gene (T13347) and protein are	
CC	useful for diagnosis or treatment of malignant tumours associated with	
CC	activation of the ras gene.	
CC	Sequence 1077 AA;	
DB	834 GTHHFSHEIAEQTLTDLAEFLYKI-DIPEVL-LMAKEQNEKSPNLTQTEHFNNMSY 891	
OY	604 GTFELMSSKDLAYQDTITDWEFLFCVHLELELYTFGHNPKKTTANDLFLRFNFNIQF 663	
DB	892 WRSITIMLOEKAQDERLLKFIKIMKHLRLNFNNSLALISLSDSAPIRLE--WOKQ 949	
OY	664 WYVETICLSQSLSKRVQLKKFKIKIAHCKEYKINLSFALVMGLSNINAVSRALTWELK 723	
DB	950 TSE--GL-AEYCTLLDSSSFRAVRAALSEVEPCIPYGLIIDLTFVHLGNDPYDGK 1006	
OY	724 PSKFKFYAEFEESLMDPSRNHRAVRLVAKLEPLIPMLIKDMFTFHGNTFIDNL 783	
DB	1007 VNEFKRMQFNILDSMRFCQAHYD 1031	
OY	784 VNEFKRMIAVTAARTVRYRSQPFN 808	
RESULT	2	

ID	RS0991	standard; protein; 489 AA.
AC	RS0991;	
DT	11-MAY-1994	(first entry)
DE	Peptide which modulates GDP exchange in the p21-GDP complex.	
KW	GDP; exchange; GRF; GDP exchange factor; antisense; diagnosis;	
KW	detection; cancer; ras; oncogene.	
OS	Homo sapiens.	
PN	MO9321314-A.	
PD	28-OCT-1993.	
PF	19-APR-1993; F00382.	
PR	21-APR-1992; FR-004827.	
PA	(RHON) RHONE POULENC RORER SA.	
PI	Schweighoffer F, Tocque B;	
DR	WPI: 93-351724/44.	
DR	N-PSDB: 051233.	
PT	New peptide(s) modulating GDP exchange in complexes - with ras	
PT	protein and derived antibodies, nucleic acid etc. esp. for	
PT	diagnosis and treatment of cancer	
PS	Claim 2; Page 25-26; 46pp; French.	
CC	The peptide antagonises the interaction of GDP-exchange factor (GRF)	
CC	with the p21-GDP complex and thereby regulates the activity of ras	
CC	gene products. It can also be used to identify other compounds	
CC	which can modulate GDP exchange. The coding sequence can be used	
CC	to produce antisense products which can inhibit oncogene expression	
CC	and for use in diagnosis e.g. detecting (over)expression of GRF for	
CC	typing of cancers.	
CC	Sequence 489 AA;	

Query Match	4.68;	Score 287;	DB 1;	Length 489;
Best Local Similarity	25.68;	Pred. No. 9.88e-13;		
Matches	40;	Conservative	49;	Mismatches 66; Indels 1; Gaps 1

```

Db 300 KHFNDISLWSEIIRNEDIARSALEKWAAYDJCRCLHNTAVLETTSSNMKRSIAFR 359
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 656 RRFNEIOFWVATEICLOSQJSLRQOLLKFKFIKAHCEKYNNLSPFAIWMGLNIAVSR 715
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 360 LKTHMLKRSKOTKALIDKQLVSESEGRFKMLBALNCOPPCVPYLGMTLTDLAFIEEG 419
      |:::| ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 716 LALTWELKPSKFKFFYAEFESLIMOPSRNHRYRLTVAKLEPPLIFEMBLIKDWTFTHEG 775
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 420 TPNTYDEGLVNSKRMMSHIIIRIQOQQAUKIE 455
      ::::: ||||| ::::: |:::| :::::
Qy 776 NKTETIDN-LVNFEXKMRMLANTARIVRYRYSOPFMD 810
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

RESULT	3
ID	R50990 standard; Protein; 666 AA.

DT 11-MAY-1994 (first entry)
DE Peptide which modulates GDP exchange in the p21-GDP complex.
KW GDP; exchange; GRF; GDP exchange factor; antisense; diagnosis;
KW detection; cancer; ras; oncogene.
OS Homo sapiens.
PN M09321314-A.
PD 28-OCT-1993.
PF 19-APR-1993; F00382.
PR 21-APR-1992; FR-004827.
PA (RHON) RHONE-POULENC RORER SA.
PI Schweighoffer F, Tocque B;
DR WPI: 93-351724/44.
DR N-PSDB: Q51233.
PT New peptide(s) modulating GDP exchange in complexes - with ras
PT protein and derived antibodies, nucleic acid etc. esp. for
PT diagnosis and treatment of cancer
PS Claim 2: Page 23-24; 46pp; French.
CC The peptide antagonises the interaction of GDP-exchange factor (GRF)
CC with the p21-GDP complex and thereby regulates the activity of ras
CC gene products. It can also be used to identify other compounds
CC which can modulate GDP exchange. The coding sequence can be used
CC to produce antisense products which can inhibit oncogene expression
CC and for use in diagnosis e.g. detecting (over)expression of GRF for
CC typing of cancers.
Q0 Sequence 666 AA;

Query Match	4.6%;	Score 287;	DB 1;	Length 666;
Best Local Similarity	25.6%;	Pred. No. 9.88e-13;		
Matches	40;	Conservative	66;	Indels 1; Gaps 1;

```

Db 477 KHHDISNLASEIIRNEDINARSALEKKVAAVADJCRCHANNVALEITSSMNRSAIIP 536
    ::||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 656 RRFNEIDPFWVTEELCICSOJSTRQULLKFKRIKIAHCKEYKRNLSFFAIVMGJNSIAVSR 715
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 537 LKKWLVSKQTKALIDKLOKLVSEGRFNLRREALKNCPCPVYGMVLTPLAFIEEG 596
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 716 LALTWELKPSFKFEKYEAFESLNDPSNNHAAVRIYAKLEPULPEWPLIKDMFTTBEG 775
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 597 TPNTYEDGLVNSFKRMISHIREIIRFOOTAIVKE 632
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 776 NKTIFDNLVNEFKRMIAINARIVRIYRSQRPND 810
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db 597 TPNYTEGLVNFESKMRMISHIIREIRQFOQTAYKIE 632

QY 776 NKTEIDN-LVNEFKRMIA NTARTVRYKRSQFNPD 810

RESULT	4
ID	R43578 standard; Protein; 814 AA.

DT	11-MAY-1994 (first entry)
DE	peptide which modulates GT

KW detection: cancer: ras: oncogene.
 KW GDP: exchange: GEF: GDP exchange factor; antisense; diagnosis;
 KW GDP: exchange: GEF: GDP exchange factor; antisense; diagnosis;

OS Homo sapiens.
PN W09321314-A.

PF 19-APR-1993; F00382.

PA (RHON) RHONE POULENC RORER SA.

DR WPI; 93-351724/44

PT New peptide(s) modulating GDP exchange in complexes - with ras protein and derived antibodies nucleic acid etc. assn for PT

PS claim 2: Page 20-22: 46bp: French.
diagnosis and treatment of cancer

CC with the p21-GDP complex and thereby regulates the activity of ras

CC which can modulate GDP exchange. The coding sequence can be used

CC and for use in diagnosis e.g. detecting (over)expression of GRF for

Sequence	814 AA
50	
100	
150	
200	
250	
300	
350	
400	
450	
500	
550	
600	
650	
700	
750	
800	
850	
900	
950	
1000	

Query Match	4.6%;	Score 287;	DB 1;	Length 814;
Best Local Similarity	25.6%;	Pred. No. 9.88e-13;		
Matches	40;	Conservative	49;	Mismatches 66;
				Indels 1;
				Gaps 1;

```
Db 625 KHEKDINSLASEIRRNEDINARVSALEKWAADICRLHNVAVELETSSMNRSAIFR 684A
::||:| : :| :: | : :| :| : :| :| : :| :| : :| :| : :| :| : :| :|
Qy 656 RRENEIOEWVTETELCSQLSKRVOLKKFIKIAHCKEYKNLSFFAIMGLSNIIVSR 715A
```

```
Db      685 LKKTWLKVSKQIKALIDKQLVSSSEGRFKNLREALRNCBPPCVPIGMYLTDLAFIEEG /44
```

02 / 10 UNIMWENEFSENF IAEFESLWDFSKNNKAI KLI VANNEFFLIRFMPLELINDMIF IHEB / 13

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

RESULT 5

AC R43580;
DE 11-MAY-1964 (Est. Est. Est.)

Peptide which modulates GDP exchange in the p21-GDP complex.

KW detection; cancer; ras; oncogene.
OS Homo sapiens

PN W09321314-A.

PD 28-OCT-1993 .
PF 19-APR-1993 : F00382 .
PR 21-APR-1992 : FR-004827 .
PA (RHON) RHONE POULENC ROREX SA .
PI Schweißhoffer F, Tocchio B,
DR WPI: 93-351724/44 .
DR N-PDB: 051235 .
PT New peptide(s) modulating GDP exchange in complexes - with ras
protein and derived antibodies, nucleic acid etc. esp. for
diagnosis and treatment of cancer
PS Claim 2: Page 32-34: 46pp: French.
CC The peptide antagonises the interaction of GDP-exchange factor (GRF)
CC with the p21 GTP complex and thereby regulates the activity of ras
CC gene products. It can also be used to identify other compounds
CC which can modulate GDP exchange. The coding sequence can be used
CC to produce antisense products which can inhibit oncogene expression
CC and for use in diagnosis e.g. detecting (over)expression of GRF for
CC typing of cancers.
CC Sequence 652 AA:
SQ

Query Match	4.48;	Score 276;	DB 1;	Length 652;
Best Local Similarity	25.8%;	Pred. No. 7.60e-12;		
Matches 59;	Conservative 71;	Mismatches 89;	Indels 10;	Gaps 7

[illegible]

RESULT 6
ID RA7048 standard; Protein; 1297 AA.
AC RA7048;
DT 26-MAY-1994 (first entry)
DE Mammalian son of sevenless (mSos 2) gene product.
KW Son of sevenless gene; mSos; Sos; guanine exchange factor; Ras;
KM tyrosine kinase; tumour.
OS Mus musculus.
PN M09316179-A.
PD 19-AUG-1993.
PF 17-FEB-1993; A00068.
PR 17-FEB-1993; A0-000921.
PT (FLOR-) FLOREY INST EXPERIMENTAL PHYSIOLOGY.
PI Bowtell DDL;
WP1: 93-272883/34.
DR N-PSDB: Q58624.
PT DNA encoding mammalian son of sevenless gene and proteins
PT encoded by it - regulates tyrosine kinase signalling pathway via
PT ras proteins, may be used for diagnosis and therapy of genetic
PT defects
PS Claim 6: Figure 2; 78pp; English.
CC The mammalian son of sevenless gene has been found to have homology
CC with various guanine exchange factors, suggesting a possible role in
CC protein tyrosine kinase pathways via the Ras proteins. Mutations in
CC the genes and consequently their products may result in specific
CC defects or tumour formation. Antagonists to the protein/gene may
CC be used to treat conditions arising from mutations.
SQ Sequence 1297 AA;

Query Match	4.38;	Score 266;	DB 1;	Length 1297;
Best Local Similarity	24.98;	Pred. No. 4.81e-11;		
Matches	57;	Conservative	70;	Mismatches 92;
			Indels 10;	Gaps 7

[illegible]

RESULT	7
ID	R84639 standard; Protein; 1297 AA.

DT	25-FEB-1996	(first entry)
DE	msos2 protein.	
KW	msos2; SOS2; Grb2; BCR-ABL; tyrosine kinase; transformation; Ras;	
KW	oncoprotein; leukaemia; Ras guanine nucleotide releasing protein.	
OS	Mus musculus.	
FX	Key	Location/Qualifiers
FT	region	1107..1119
FT	region	/label=Proline-rich_motif
FT	region	1133..1145
FT	region	/label=Proline-rich_motif
FT	region	1164..1176
FT	region	/label=Proline-rich_motif
FT	region	1178..1189
FT	region	/label=Proline-rich_motif
PM	CA2113494-A.	
PD	15-JUL-1995.	
PF	14-JAN-1994; 113494.	
PR	14-JAN-1994; CA-113494.	
RA	(MOUN) MOUNT SINAI HOSPITAL CORP.	
RA	(TEXA) UNIV TEXAS.	
PI	Artinghaus R, Gish G, Liu J, Pawson A, Pull L;	
PI	WPI; 95-302931/40.	
DR	N-FSDB; T05114.	
PT	Detection of agents that modify BCR-ABL mediated transformation -	
PT	useful in treatment of leukaemia and other malignancies	
PS	Disclosure; Page 62-67; 10ppp; English.	
CC	Mouse msos2 (R84639) is a Ras guanine nucleotide releasing protein.	
CC	The SH3 domains of Grb2 associate the the proline-rich domains of	
CC	msos2 and msos1 (R84638). Synthetic peptides (R84640-48) based on	
CC	these motifs can be used to screen for compounds that affect BCL-ABL	
CC	mediated transformation. Such compounds have value in the treatment	
CC	of chronic, acute myelogenous or acute lymphocytic leukaemia.	
SQ	Sequence 1297 AA;	

Query Match	4.28;	Score 262;	DB 1;	Length 1297;
Best Local Similarity	24.98;	Pred. No. 1.00e-10;		
Matches 57; Conservative		70;	Mismatches 92;	Indels 10; Gaps 7;

[illegible]

QY 763 ELLIKDMFTFHGKNTFI---D-RLVNFEMKRMANTARTVRYRSQPF 807

RESULT 8
R43579 standard; Protein; 364 AA.
AC R43579;
DE 11-MAY-1994 (first entry)
KW Peptide which modulates GDP exchange in the p21-GDP complex.
KM GDP; exchange; GRF; GDP exchange factor; antisense; diagnosis;
KW detection; cancer; ras; oncogene.
OS Homo sapiens.
PN WO9321314-A.
PD 28-OCT-1993.
PF 19-APR-1993; F00382.
PR 21-APR-1992; FR-004827.
PA (RHON) RHONE POULENC ROBER SA.
PI Schweihofer F, Tocque B;
DR N-PSDB; Q51234.
PT New peptide(s) modulating GDP exchange in complexes - with ras
PT protein and derived antibodies, nucleic acid etc. esp. for
PT diagnosis and treatment of cancer
PS Claim 29: Page 28-29; 46pp; French.
CC The peptide antagonises the interaction of GDP-exchange factor (GRF)
CC with the p21-GDP complex and thereby regulates the activity of ras
CC gene products. It can also be used to identify other compounds
CC which can modulate GDP exchange. The coding sequence can be used
CC to produce antisense products which can inhibit oncogene expression
CC and for use in diagnosis e.g. detecting (over)expression of GRF for
CC typing of cancers.
SQ Sequence 364 AA;

Query Match 4.1%; Score 256; DB 1; Length 364;
Best Local Similarity 25.4%; Pred. No. 3.02e-10;
Matches 58; Conservative 65; Mismatches 97; Indels 8; Gaps 6;

Db 22 FOSSPTVEWHISRPGHETFDLTLHPTEIARQLTLDSDIRAVOPSLVGSVWTKRD 81
QY 586 FDSLTPLPQEGEPYGVGTGFELMS--SKDLAYQMTYIDWELFNCHELELIYHTFGRRN 643
Db 82 KEINSPNLKMTIRHTNTLTFEKCIVETENLEERAVAVSRHIELLOFQELNENGYLE 141
QY 644 FKKTANLDFLRFRNEIOFWVWEICLSQLSRQVOLLKFKIKIAHCKEYKLNLSFPA 703
Db 142 VVSAMSSPVYRLDHTFPQIPSRQKILIEAHELSE-DHYKK-YLAKRSINPCVPFPG 199
QY 704 IWMGLSNIAVSKLALTWELPSKRFKFAEFESLMDSRNRRARLTVAKLEPPLIPFMP 763
Db 200 IYLNILKTEGPNPEVLKRRHKGELINFSKRRVAETGEIOQYONOPY 247
QY 764 LLIKDMFTFHGKNTFI---D-RLVNFEMKRMANTARTVRYRSQPF 807

RESULT 9
R47043 standard; Protein; 1319 AA.
AC R47043;
DE 26-MAY-1994 (first entry)
KW Mammalian son of sevenless (mos 1) gene product.
KM Son of sevenless gene; mos; Sos; guanine exchange factor; Ras;
KW tyrosine kinase; tumour.
OS Mus musculus.
PN WO9316179-A.
PD 19-AUG-1993.
PF 17-FEB-1993; AU00068.
PR 17-FEB-1992; AU-000921.
PA (FLOR) FLOREY INST EXPERIMENTAL PHYSIOLOGY.
PI Bowtell DDL;
DR MPI: 93-272883/34.
DR N-PSDB; Q48610.
PT DNA encoding mammalian son of sevenless gene and proteins
PT encoded by it - regulates tyrosine kinase signalling pathway via
PT ras proteins, may be used for diagnosis and therapy of genetic
PT defects

PS Claim 4; Figure 1; 78pp; English.
CC The mammalian son of sevenless gene has been found to have homology
CC with various guanine exchange factors, suggesting a possible role in
CC protein tyrosine kinase pathways via the Ras proteins. Mutations in
CC the genes and consequently their products may result in specific
CC defects or tumour formation. Antagonists to the protein/gene may be
CC used to treat conditions arising from mutations.
SQ Sequence 1319 AA;

Query Match 4.1%; Score 255; DB 1; Length 1319;
Best Local Similarity 25.4%; Pred. No. 3.63e-10;
Matches 58; Conservative 64; Mismatches 98; Indels 8; Gaps 6;

Db 754 FOSSPTVEWHISRPGHETFDLTLHPTEIARQLTLDSDIRAVOPSLVGSVWTKRD 813
QY 586 FDSLTPLPQEGEPYGVGTGFELMS--SKDLAYQMTYIDWELFNCHELELIYHTFGRRN 643
Db 814 KEINSPNLKMTIRHTNTLTFEKCIVETENLEERAVAVSRHIELLOFQELNENGYLE 873
QY 644 FKKTANLDFLRFRNEIOFWVWEICLSQLSRQVOLLKFKIKIAHCKEYKLNLSFPA 703
Db 874 VVSAMSSPVYRLDHTFPQIPSRQKILIEAHELSE-DHYKK-YLAKRSINPCVPFPG 931
QY 704 IWMGLSNIAVSKLALTWELPSKRFKFAEFESLMDSRNRRARLTVAKLEPPLIPFMP 763
Db 932 IYTNILKTEGPNPEVLKRRHKGELINFSKRRVAETGEIOQYONOPY 979
QY 764 LLIKDMFTFHGKNTFI---D-RLVNFEMKRMANTARTVRYRSQPF 807

RESULT 10
R84638 standard; Protein; 1336 AA.
AC R84638;
DE 25-FEB-1996 (first entry)
DE mSOS1 protein.
KW mSOS1; SOS1; Grb2; BCR-ABL; tyrosine kinase; transformation; Ras;
KW oncoprotein; leukaemia; Ras guanine nucleotide releasing protein.
OS Mus musculus.
FH Key location/qualifiers
FT region 1149..1161
FT /label= proline-rich_motif
FT region 1178..1190
FT /label= proline-rich_motif
FT region 1210..1222
FT /label= proline-rich_motif
FT region 1288..1300
FT /label= proline-rich_motif
PN CA2113494-A.
PD 15-JUL-1995.
PF 14-JAN-1994; 113494.
PR 14-JAN-1994; CA-113494.
PA (MOUN) MOUNT SINAI HOSPITAL CORP.
PA (TEXA) UNIV TEXAS.
PI Arlinghaus R, Gish C, Liu J, Pawson A, Pull L;
DR MPI: 95-302931/40.
DR N-PSDB; T05113.
PT Detection of agents that modify BCR-ABL mediated transformation -
PT useful in treatment of leukaemia and other malignancies
PS Disclosure; Page 54-59; 106pp; English.
CC Mouse mSOS1 (R84638) is a Ras guanine nucleotide releasing protein
CC that combines with BCR-ABL and Grb2 (R84636) to activate the Ras
CC pathway leading to morphological transformation. The SH3 domains
CC of Grb2 associate the the proline-rich domains of mSOS1 and mSOS2
CC (R84639). Synthetic peptides (R84640-48) based on these motifs
CC can be used to screen for compounds that affect BCR-ABL mediated
CC transformation. Such compounds have value in the treatment of
CC chronic, acute myelogenous or acute lymphocytic leukaemia.
SQ Sequence 1336 AA;

Query Match 4.1%; Score 255; DB 1; Length 1336;
Best Local Similarity 25.4%; Pred. No. 3.63e-10;
Matches 58; Conservative 64; Mismatches 98; Indels 8; Gaps 6;

Db 771 POSPPTVEMHHSRGIETFDLTLHPRIAROLTLLESDYRANOPSELVSVTKED 830
 Oy 586 FDSLPDQEGPVTGIGTELMS--SKDIATOMTIDWELFCHVHELITHTFGRRH 643
 Db 831 KEINSPNLMKIRHTNTLTFEKCIVETENLIERVAVSRITIELLOFOLNFGVLE 890
 Oy 644 FKTTANLDLFRBNFEIOWVWVEICLCSQLSKRVOLKFKIRIAHCKEYKLNLSFA 703
 Db 891 VVSANSSPVYRLDHTFQIPSRKQKILEFHEHSEF-DHKKK-YLAKLSINPCVPEFG 948
 Oy 704 IYMGSLNIAVSRLATLTKLPSKFKKFAEFESLMDPSRNRARLTVALEPLLPFMD 763
 Db 949 IYLTILKTEGNEPEVLRHKKELINESKRRRAVETGEIQOYONOPY 996
 Oy 764 LLIKMTFTHEGN-KTFI--DN-LVNEFKMKMIANTARTVRYRSQPF 807

RESULT 11
 ID W54350 standard; protein; 380 AA.
 AC W54350.
 DT 14-MUG-1998 (first entry)
 DE CAMP-dependent protein kinase type I-beta regulatory chain.
 KM Endometrium; hyperplasia; adenocarcinoma; proliferative phase;
 OS 2D gel electrophoresis; detection.
 PN Homo sapiens.
 PD W09810291-A1.
 PF 05-SEP-1997; G02394.
 PR 08-APR-1997; GB-007132.
 PR 06-SEP-1996; GB-018600.
 PA (CLIN-) CENT CLINICAL & BASIC RES.
 PI Bystalsen I, Rev SJ, Larsen P.
 DR WPI:98-207057/18.
 PT Biochemical markers of human endometrium - useful for, e.g.
 PT diagnosis of hyperplasia and adenocarcinoma
 PS Disclosure; Page 19-20; 77pp; English.
 CC Proteins W54349-W54364 are examples of proteins produced in the
 CC endometrium during the hyperplasia, adenocarcinoma or proliferative
 CC phase of the endometrium. The presence and quantities of these proteins
 CC can be detected using 2D gel electrophoresis comparison of cell lysates.
 CC The proteins can be used as biochemical markers to detect the phase of
 CC the endometrium and can be measured in body fluids, obviating the need
 CC for endometrial biopsies.
 SQ Sequence 380 AA;

Query Match 3.2%; Score 201; DB 1; Length 380;
 Best Local Similarity 31.7%; Pred. No. 5.94e-06;
 Matches 32; Conservative 26; Mismatches 40; Indels 3; Gaps 3;
 Db 137 FAHLDNERSDIFDAMPVTHIAGEFTV-QQNGNEGNEFVVYVNGEWT-NIS 194
 Oy 195 LSHLSTYVRELAGVILFESHKGGTVLFDNGEGTSTYILIKGSVNVVYIGKGVCTLH 254
 Db 195 EGGSGFELALYCTPRRAVAKTDL-KLWGIDDSYRIL 234
 Oy 255 EGGDFGKALVNDAPRAAIVREDNCHFLVYDKEDENRIL 295

RESULT 12
 ID W87995 standard; protein; 609 AA.
 AC W87995;
 DT 15-APR-1999 (first entry)
 DE An alternatively spliced human MCG7 protein.
 KM MCG4 protein; gene regulatory function; heat shock protein;
 KM guanine nucleotide exchange factor protein; MCG7 protein;
 KM heat shock-binding protein; MCG18 protein; zinc finger protein;
 OS Homo sapiens.
 PN W09853061-A1.
 PD 26-NOV-1998.
 PF 22-MAY-1998; AU0380.
 PR 22-JAN-1998; AU-001460.
 PR 23-MAY-1997; AU-006972.

PR 23-MAY-1997; AU-006973.
 PR 23-MAY-1997; AU-006974.
 PR 22-JAN-1998; AU-001458.
 PR 22-JAN-1998; AU-001459.
 PA (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
 PI Gartside M, Grimmond S, Hancock J, Hayward N, Silins G;
 DR WPI:99-070146/06.
 DR N-PSDB: X04553.
 PT New gene-expression regulatory genes, mcg4, mcg7, and mcg18 - encode
 PT a zinc finger protein, a GEF, and a heat shock or heat shock binding
 PT protein, useful to detect and treat cancer
 PS Claim 5; Fig 13b; 80pp; English.
 CC The present sequence represents a MCG7 protein. The protein has gene
 CC regulatory functions, and has homology to a heat shock protein or
 CC heat shock-binding protein. The specification also describes MCG4,
 CC which is homologous to a guanine nucleotide exchange factor protein,
 CC and MCG18, which is homologous to a zinc finger protein.
 CC Detection of mutations in the MCG genes can be used to identify the
 CC propensity for various types of cancer, and to treat, arrest, or
 SQ otherwise ameliorate, the effects of a cancer in an animal or bird.
 Sequence 609 AA;

Query Match 2.9%; Score 180; DB 1; Length 609;
 Best Local Similarity 26.4%; Pred. No. 2.29e-04;
 Matches 38; Conservative 34; Mismatches 68; Indels 4; Gaps 4;
 Db 150 FDHPEMELAHLYLEYRSF-C-KILFODYHSFVTHGCTVDPVLERFISLNSYSONV 207
 Oy 606 FELSSKDLAYOMTIDWELFCHVHELITHTFGRRHFKTTANLDLFRFNEIOFWV 665
 Db 208 QLMT-LSKPTAPORALVTHFVHAETKLOLONPNTLMVYVGGTSSHSRLKEPHSHS 266
 Oy 666 VTEICLSQLS-KRYQLKFKIRIAHCKEYKLNLSFRAIVMGLSNIAVSRLATWELKP 724
 Db 267 PETIKMEGLNELVATGATNGYNR 290
 Oy 725 SKKFFVAFESLMDPSRNRHAYR 748

RESULT 13
 ID W87994 standard; protein; 728 AA.
 AC W87994;
 DT 15-APR-1999 (first entry)
 DE A human MCG7 protein.
 KM MCG4 protein; gene regulatory function; heat shock protein;
 KM guanine nucleotide exchange factor protein; MCG7 protein;
 KM heat shock-binding protein; MCG18 protein; zinc finger protein;
 KM cancer.
 OS Homo sapiens.
 PN W09853061-A1.
 PD 26-NOV-1998.
 PF 22-MAY-1998; AU0380.
 PR 22-JAN-1998; AU-001460.
 PR 23-MAY-1997; AU-006972.
 PR 23-MAY-1997; AU-006973.
 PR 23-MAY-1997; AU-006974.
 PR 22-JAN-1998; AU-001458.
 PR 22-JAN-1998; AU-001459.
 PA (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
 PI Gartside M, Grimmond S, Hancock J, Hayward N, Silins G;
 DR WPI:99-070146/06.
 DR N-PSDB: X04552.
 PT New gene-expression regulatory genes, mcg4, mcg7, and mcg18 - encode
 PT a zinc finger protein, a GEF, and a heat shock or heat shock binding
 PT protein, useful to detect and treat cancer
 PS Claim 5; Fig 13a; 80pp; English.
 CC The present sequence represents a MCG7 protein. The protein has gene
 CC regulatory functions, and has homology to a heat shock protein or
 CC heat shock-binding protein. The specification also describes MCG4,
 CC which is homologous to guanine nucleotide exchange factor protein,
 CC and MCG18, which is homologous to a zinc finger protein.
 CC Detection of mutations in the MCG genes can be used to identify the
 CC propensity for various types of cancer, and to treat, arrest, or

